

70427

Access DB#

192

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: RITAM MITRA Examiner #: 77995 Date: 7/8/02
 Art Unit: 1653 Phone Number 30605-1211 Serial Number: 09/691344
 Mail Box and Bldg/Room Location: 9B01/EM1 Results Format Preferred (circle): PAPER DISK E-MAIL
Ln 9B.03

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: "Novel human proteins and polynucleotides encoding the same"

Inventors (please provide full names): Gregory Donoho, C. Alexander Turner, Michael Nchla, Glenn Frie-
drich, Brian Zambrowicz, Arthur Sands

Earliest Priority Filing Date: 10/19/1999

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

I would request an expedited literature (both patent and Non-patent) search on above application because this is a date case. Don't do Sequence
the search.

The search should encompass nucleotides that encode proteins, wherein proteins have structural similarity to animal CUB domain proteins, coagulation factors V and XIII, milk fat globule-EGF factor 8, transcriptional repressor AE-binding protein-1 and neuropilins 1 and 2.

Keywords: CUB domain, discordin domain, NHP, agonist, antagonist, receptor, protein

Mary Hale - Supervisor, Info. Branch
 STIC - Biotech/Chem. Library
 CM-1 Room ED1
 703-308-4258

STAFF USE ONLY

Searcher: Mary Type of Search: NA Sequence (#) STN 120,70
 Searcher Phone #: AA Sequence (#) Dialog
 Searcher Location: Structure (#) Questel/Orbit
 Date Searcher Picked Up: Bibliographic Dr.Link
 Date Completed: 7/9 Litigation Lexis/Nexis
 Searcher Prep & Review Time: Fulltext Sequence Systems
 Clerical Prep Time: Patent Family WWW/Internet
 Online Time: 18 Other Other (specify)

The sequence represents a cDNA which encodes human TANC0_229 polypeptide. This protein and similar others exhibit the ability to affect growth, proliferation, survival, differentiation, activity, morphology, or movement/migration of, e.g., T cells and cells of the heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus tissue. They can be used as modulating agents for regulating cellular processes, thus, they can be used as modulating agents for regulating cellular processes, thus, they can prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma,

[illegible]

File copy

SEA ID NO: 4

Database: A-Geneseg-032802

AC NO: AAU00670

QY 288 GSLSDRKFLFTSNCGSRSLSEFPDGOIRASSWSQVNSGDOYHNSPGCARLDDGSPMA 347
 DB 241 gslsdrrflftsnngcsrslsefpdgoirasswsqvnsgdqyhwspgqarlqdgpswa 300
 QY 348 SGSSNNHKKREWELEIDLGKKKITGIRTTGSTOSNFNFYKSFVNNFKNNNSKMTYKG 407
 DB 301 sgssnnhkprrwleldlgkkitgirtgstosnfnyksfvnnfknnnskmtykg 360
 QY 408 IVNNEKVFQNSNFRDPQNNFIPPIVARYVAVVQPMHORIAKVELIGCOITGND 467
 DB 361 ivnnekvfgnfnfrdpqnnfippivaryvavvqpmhoriakveligcoitgnds 420
 QY 468 LVWRKTSOSTSVSTKKEDEITRPPISESTSGINTTVAIPVLVLLVAVFAGMGIFA 527
 DB 421 lvwrktsostsvstkkedelitrrpipsestsginttvaipvlvllvavfagmgiffa 480
 QY 528 RKKKKKSPYGSAAEAKTDCWKQIKYPPARHQAETISYDNKEMTKLDTLITSDMA 586
 DB 481 rkkkkkspygaaeqktcdwqkikypfarhgsaetislydnkemtkldlitsdma 539
 RESULT 3
 AAU00670
 ID AAU00670 standard; Protein; 715 AA.
 AC AAU00670;
 DT 07-SEP-2001 (first entry)
 XX
 DE Human TANGO 229 polypeptide.
 XX
 KW Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anapylaxis; hepatitis; multiple sclerosis; coronary artery disease;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immune-competence; vertebrate; blood; serum.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Peptide
 FT 1..34 Location/Qualifiers
 FT /note="Signal peptide"
 FT 35..455
 FT Domain
 FT /note="Extracellular domain"
 FT 35..715
 FT Protein
 FT /note="Mature human TANGO 229"
 FT 456..480
 FT /note="Transmembrane domain"
 FT 481..715
 FT Domain
 FT /note="Cytoplasmic domain"
 XX
 PN WO200129088-A1.
 PD 26-APR-2001.
 PF 23-JUN-2000; 2000MO-US17386.
 PR 19-OCT-1999; 99US-0420707.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Mackay CR, Myers PS, Kilst SJ, Fraser CC, Leiby KR.
 DR WPI: 2001-308477/32.
 DR N-PSDB: AAS00660.
 PT New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes

XX Claim 8; Fig 1; 263pp; English.
 PS
 CC The sequence represents human TANGO 229 polypeptide. This protein and
 CC similar others exhibit the ability to affect growth, proliferation,
 CC survival, differentiation, activity, morphology, or movement/migration
 CC of, e.g. T cells and cells of the heart, liver, pancreas, placenta,
 CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
 CC blood leukocyte, bone marrow or thymus tissue. They can be used as
 CC modulating agents for regulating cellular processes, thus, the proteins
 CC and their associated nucleic acids can be used to prognosticate, prevent,
 CC diagnose, or treat disorders associated with physiological processes.
 CC These disorders include abnormal blood coagulation, asthma, anapylaxis,
 CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
 CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
 CC deficit disorder, Crohn's disease, gastroenteritis, goitre,
 CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
 CC muscular dystrophy. Antibodies to disorders such as these can be made by
 CC providing a polypeptide of the invention to an immuno-competent
 CC vertebrate and harvesting blood or serum from the vertebrate.
 XX
 S0 Sequence 715 AA:

Query Match 92.1%; Score 2812; DB 22; Length 715;
 Best Local Similarity 100.0%; Pred. No. 2.8e-243;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 MYPGARGGALARAAGGLALLIAVASAPRLQAEELGDCGHLVYQDSGTMTSKNYPG 107
 DB 1 mvparggalaraagglalliavasaprlqaeelgdcgchlvyqdsgmtsknypg 60
 QY 108 TYPNHTVCECTIYTPKRRILRLRGDIDISQTCASDYLFTSSDQYGYCGSMYVPE 167
 DB 61 typnhtvcectiytpkrrilrlrgdidisqtcasyllftssdqygycgsmvpe 120
 QY 168 LLNTSSTVTRFESSHISGRGFLTYASSDHDILTJCFERASHYKTEKSCPCAGCRD 227
 DB 121 llntsevtvrfesshigrgrfltyassdhdiltcferashykteyskpcagcrd 180
 QY 228 VAGDISGMVDGYRDSILCKAALHAGTIDELGGQISYLQKRGISRYEGILANGVLSRD 287
 DB 181 vagdisgmvdgyrdslckaalhagtidelggqisvlyqrgisryegilangvlsrd 240
 QY 288 GSLSDRKFLFTSNCGSRSLSEFPDGOIRASSWSQVNSGDOYHNSPGCARLDDGSPMA 347
 DB 241 gslsdrrflftsnngcsrslsefpdgoirasswsqvnsgdqyhwspgqarlqdgpswa 300
 QY 348 SGSSNNHKKREWELEIDLGKKKITGIRTTGSTOSNFNFYKSFVNNFKNNNSKMTYKG 407
 DB 301 sgssnnhkprrwleldlgkkitgirtgstosnfnyksfvnnfknnnskmtykg 360
 QY 408 IVNNEKVFQNSNFRDPQNNFIPPIVARYVAVVQPMHORIAKVELIGCOITGND 467
 DB 361 ivnnekvfgnfnfrdpqnnfippivaryvavvqpmhoriakveligcoitgnds 420
 QY 468 LVWRKTSOSTSVSTKKEDEITRPPISESTSGINTTVAIPVLVLLVAVFAGMGIFA 527
 DB 421 lvwrktsostsvstkkedelitrrpipsestsginttvaipvlvllvavfagmgiffa 480
 QY 528 RKKKKKSPYGSAAEAKTDCWKQIKYPPARHQAETISYDNKEMTKLDTLITSDMA 586
 DB 481 rkkkkkspygaaeqktcdwqkikypfarhgsaetislydnkemtkldlitsdma 539
 RESULT 4
 AAU00628
 ID AAU00628 standard; Protein; 487 AA.
 AC AAU00628;
 DT 29-NOV-2001 (first entry)

us-09-691-344a-3.rnq

Page 10

Qy	132	gtccccacacatgycacacacagagatagccttgaaggtgagctcatgtgtccagatt	1383
Db	139	gtccccacacatgycacacacagagatagccttgaaggtgagctcatgtgtccagatt	198
Qy	1384	accacaggtgaacgatcatctgtgtgagcaca--gacaagfcaaacgaccagt	1434
Db	199	cacaaaggaatgatcatctgtgtgagcaca--gacaagfcaaacgaccagt	250
RESULT	10		
ID	AAZ32182	standard; cDNA; 6893 BP.	
XX	AAZ32182;		
AC	AAZ32182;		
XX			
DT	13-JAN-2000	(first entry)	
XX			
DE	Human coagulation factor V nucleotide sequence.		
XX			
KW	Human; coding sequence polymorphism; vascular pathology gene;		
KM	polymorphic site; phenotype correlation; forensic; paternity testing;		
KW	medicine; genetic analysis; vascular disease; ds.		
OS	Homo sapiens.		
XX			
PN	W09950454-A2.		
XX			
PD	07-OCT-1999.		
XX			
PF	26-MAR-1999; 99MO-U506473.		
XX			
PR	01-APR-1998; 98US-0054272.		
XX			
PA	(WHEED) WHITEHEAD INST BIOMEDICAL RES.		
XX			
PI	Lander ES, Daley GQ, Cargill M, Ireland JS, Rozen SG;		
XX			
DR	WPI: 1999-620066/53.		
XX	P-PSDB; AAY49363.		
PT	Determination of polymorphisms in genes, especially those identifying		
XX	predisposition to vascular disease		
PS	Claim 1; Fig 26; 134pp; English.		
XX			
CC	AAZ32159 to AAZ32194 represent reference alleles for specifically		
CC	claimed nucleic acid sequences from the present invention which comprise		
CC	polymorphic sites as given in a table in the specification, selected		
CC	from 92 single nucleotide polymorphisms in which the nucleotide at the		
CC	polymorphic site is different from a nucleotide at the same site in a		
CC	reference allele. The nucleic acids, and primers and probes, are used to		
CC	identify polymorphisms, which may predispose an individual to disease,		
CC	especially a vascular disease. They can also be used in phenotype		
CC	correlations, forensics, paternity testing, medicine or genetic		
CC	analysis. AAY49350 to AAY49573 represent the proteins which correspond		
CC	to some of the reference alleles.		
Q	Sequence 6893 BP; 2090 A; 1700 C; 1423 G; 1680 T; 0 other.		
XX			

Db 6396 aatccccgagcgtgtgaatgctcgtgcaagcgaacacacataagc-----ag 6439
 Qy 1078 tggcggagatcgtgttg999gagaaagaaataaagaattaggaccagatc 1137
 Db 6440 tggctagaatctactcaagatcaagaataacgcaattatacacaaggtcgc 6499
 Qy 1138 acacagtcgaactcaacttattatgttaagatttggatgaactcaaaaataat 1197
 Db 6500 aagctcgtcctcctgaatgtatgttaagcctataccactacagtgagcaggga 6559
 Qy 1198 tctaagtggaagaccataaaggaattgtgaataatgaagaagaagtggtcaaggaac 1257
 Db 6560 gtggaatggaacacatacgaagctgaatccctcctcgtggaagaagattttgaaggaat 6619
 Qy 1258 tctaacttcggagcccgatgcaaaacattcctccctcctcgtggaagataatgtg 1317
 Db 6620 actataccaagaacatgtgaagaacttttccaacccccaacattccaaggtttac 6679
 Qy 1318 cgggtgtcccccagacatggaacagaggaatgacttgaagtgagatcattgttc 1377
 Db 6680 cgtgtcctccctaaacatggaatcaagattgctcgtcgtggaactcttggctgt 6739
 Qy 1378 cagatt 1383
 Db 6740 gatatt 6745

RESULT 11

ID AAT03920 standard; cDNA; 6909 BP.
 AC AAT03920;

DT 20-DEC-1995 (first entry)
 XX

DE Human Factor-V wild-type cDNA.
 XX

KW Factor-V; thrombosis; thrombophilia; diagnosis; anticoagulant;
 activated protein-C; APC; homozygosity; heterozygosity; ss.
 XX

OS Homo sapiens.
 XX

PN W09521938-A1.
 XX

PD 17-AUG-1995.
 XX

PF 14-FEB-1995; 95MO-XP00553.
 XX

PR 14-FEB-1994; 94EP-0200377.
 XX

PA (UYLE-) RIJMSUNIV LEIDEN.
 XX

PI Bertina RM, Reltsma PH;
 XX

DR WPI; 1995-293134/38.
 XX

PT Screening for genetic defect associated with thrombosis and/or poor
 PT anticoagulant response to activated protein C - useful to determine
 PT homozygosity or heterozygosity for a mutation in Factor V, Va, VIII
 PT or VIIa.
 XX

PS Claim 16; Page 60-65; 98pp; English.
 XX

CC The probes/primers given in AAT03921-31 are used to isolate and/or
 CC amplify and/or detect mutations in human Factor-V DNA (AAT03920),
 CC esp. in DNA encoding the APC binding and/or cleavage site.
 CC
 XX

SO Sequence 6909 BP; 2096 A; 1700 C; 1430 G; 1683 T; 0 other;
 XX

Query Match 5.3%; Score 92.8; DB 16; Length 6909;
 Best Local Similarity 51.0%; Pred. No. 2.3e-13;

Matches 248; Conservative 0; Mismatches 232; Indels 6; Gaps 1;
 Qy 898 aatgttgacagcagatccttgaatttttgaacctgacgggcaatcagagcttccca 957
 Db 6280 atctgagcttcacacacccctggtatggaatggaagatagaacaagcaatcac 6339
 Qy 958 tggcagtcgctaatgagatgagacccaagttcgtctctcgtgccaagccgact 1017
 Db 6340 gcttctcgttttaagaatcttgggtggagattacttgggaaccttcgtgctgtg 6399
 Qy 1018 caggacccaagcccatcattgtgctcgggcagatgacacacacacacacagag 1077
 Db 6400 aatgcccaaggagcgttgaatgtcctggaagcgaagcaacacacataagc-----ag 6453
 Qy 1078 tggctggaatctgatttgg999gagaaagaaataaagaagattgagccacagatc 1137
 Db 6454 tggctgaatctgactactcaagacataagaataacgcaattatacacagggtcgc 6513
 Qy 1138 acacagtcgaacttcaacttttattgaagatttggatgaactcaaaaataat 1197
 Db 6514 aagctcgtcctcctgaatgtatgttaagagctataccactacacagtgagcaggga 6573
 Qy 1198 tctaagtggaagaccataaaggaattgtgaataatgaagaagaagtggttcaaggaac 1257
 Db 6574 gtggaatggaacacatacagaagctgaatactccatgtgtgacagaattttgaaggaat 6633
 Qy 1258 tctaacttcggagcccgatgcaaaacatttattccctccctcgtggaagataatgtg 1317
 Db 6634 actaaataccaagaacatgtgaagaacttttccaacccccaacttccaggtttac 6693
 Qy 1318 cgggtgtcccccagacatggaacagaggaatgacttgaagtgagatcattgttc 1377
 Db 6694 cgtgtcctccctaaacatggaatcaagaatgaatatacctcgcctggaactcttggctgt 6753
 Qy 1378 cagatt 1383
 Db 6754 gatatt 6759

RESULT 12

ID AAA60448 standard; cDNA; 6909 BP.
 AC AAA60448;

DT 09-OCT-2000 (first entry)
 XX

DE Human factor V cDNA sequence SEQ ID NO:26.
 XX

KW Factor V; FV; activated protein C; APC; anticoagulant;
 activated protein C resistant factor V; thrombosis; screening;
 thrombophilia; ds.
 XX

OS Homo sapiens.
 XX

PN US606778-A.
 XX

PD 23-MAY-2000.
 XX

PF 06-NOV-1996; 96US-0746111.
 XX

PR 06-NOV-1996; 96US-0746111.
 XX

PA (UNMI) UNIV MICHIGAN.
 XX

PI Ginsburg D, Cui J;
 XX

DR WPI; 2000-410682/35.
 XX

PT New transgenic mice expressing activated protein C resistant factor V
 PT and factor V null transgenic mice useful for screening anticoagulants,
 PT as models for human thrombophilia and as models for testing in utero
 PT gene therapy protocols -

1

```
Query Match      11.6%;  Score 333.3;  DB 21;  Length 805;
```

```
Query Match      11.6%;  Score 353.5;  DB 21;  Length 889;
```

```
Query Match      11.6%;  Score 353.5;  DB 21;  Length 889;
```


Best Local Similarity 23.1%; Pred. No. 2.3e-22;
Matches 120; Conservative 73; Mismatches 189; Indels 137; Gaps 17;

QY 49 VPGRRGGGALARRAGGLALLAVSNAPLRLQAEEIGD-GCGHLVYVQSGTMTSKNPG 107
DB 15 vpsgtgdaa-----qparravrsqdpccggrnkskaaglytspypq 57
QY 108 TYPHATVCEKTIYVPR-KRLLIRLG-DLIESQTCASDYL-SSSDQVGPYCGS 161
DB 58 dypshqgcwvlyvapepnqkivlnfnphfelekhdkygfelfrdgsesadllgkhcgn 117
QY 162 MTPVKELLMTSEVTFRESGSHISGRFLTY-ASSP----- 198
DB 118 I-aptlissgsvlyikfidsdyarqagfsllyelftkgsedcsknfsngtlespyp 176
QY 199 --HPDLITC-----LEBASHTKTEYSK-----CPACRVDAGDIS--GNMV 237
DB 177 ekypnhdctftllakprmeillqflltdlehpdlqvagsgdkywdldwglpvpqll 236
QY 238 DGYRDTSLCKAIIHAGIADLEGQISVLQKRGISRY-----EG 277
DB 237 gkygcgktpsklrstqglstlftldmavakdgfsaryyllhpeppenfqcnpvgmesg 296
QY 278 ILANGVLSRDGLSDKRFLFTSNGCSRLSFPDQIRASSWQSVNCSQVHMSPGQA 337
DB 297 rlanegiasstfsdgr-----wtpqqs 319
QY 338 RLDDQSPNASGDSNNHPRKMLEIDGKKRTIGRTTG--STQSNFNFVYVSNFVNF 395
DB 320 rllgddngwcpnldsn---keylgvdrlflmltalatqalsretkgyvyskylev 375
QY 396 KNNNSKMTYKGIYVNEEFQGNFNFRPQNNFPIVARYVAVVQTHQIALKVE 455
DB 376 stngedkmmyrgrkh--kifamndatevlnklmpdltrfltrpqlwhglalrle 433
QY 456 LIGCOITGNDSLVWRKTS--QSTSVTKKEDETITRP 491
DB 434 lfgrcvtdapcsnmlgmisgladtqiasastreyllwsp 472

RESULT 9
AAB6308
ID AAB6308 standard; Protein; 925 AA.
AC AAB6308;
XX
XX
XX 28 JUN-1999 (first entry)
DE Neuropilin-2.
XX
XX Neuropilin; neuropilin-2; semaphorin; transmembrane protein; axon;
KM neuron; development; cell growth; immune response;
KM viral pathogenesis; treatment; disease; graft rejection;
KM viral disease; oncological disease; cancer; screening; probes.
OS Rattus rattus.
XX
XX W09904263-AT.
XX 28 JAN-1999.
XX PD 17 JUL-1998; 98MO-US14632.
XX PF 17 JUL-1997; 9YUS-0052762.
XX PR 17 JUL-1997; 9YUS-0052762.
XX
XX (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX PA Glutyl DD, Kolodkin AL.
XX PI WPI: 1999-132446/11.
XX DR N-PSDB: AAX08417.
XX PT New isolated semaphorin receptor, neuropilin-2 - used to develop

PT products for the diagnosis and treatment of neurological,
XX immunological, oncological and viral diseases
XX

PS Claim 13; Page 77-79; 88pp; English.

XX The neuropilins are type I transmembrane proteins and act as
CC semaphorin III (Sema III) receptors. The semaphorins have been
CC shown to function in repulsive axon guidance. Sema III is a
CC secreted protein that in vitro causes neuronal growth cone
CC collapse and chemorepulsion on neurites and is required in vivo for
CC correct sensory afferent innervation and other aspects of
CC development. Agents which inhibit or enhance the interaction of a
CC semaphorin and a neuropilin can be potent modulators of nerve cell
CC growth, immune responsiveness, and viral pathogenesis, and can be
CC used in the treatment and diagnosis of neurological disease.
CC neuro-regeneration, immune modulation including hypersensitivity and
CC graft-rejection, and diagnosis and treatment of viral and oncological
CC infectious diseases. The neuropilins, neuropilin-encoding nucleic
CC acids, and unique portions also are useful in screening chemical
CC libraries for regulators of semaphorin-mediated cell activity, and in
CC genetic mapping as probes for related genes, as diagnostic reagents
CC for genetic, neurological, immunological and oncological disease.

SQ Sequence 925 AA:

Query Match 11.5%; Score 351.5; DB 20; Length 925;
Best Local Similarity 24.4%; Pred. No. 3.7e-22;
Matches 114; Conservative 76; Mismatches 181; Indels 97; Gaps 17;

QY 88 CSHLYVYDQSGTMTKNTGYPYNTVCEKTIYVPR-KRLLIRLG-DLIESQTCASDY 145
DB 28 csgrlnskdaglytspypqdyshqgcwvlyvapepnqkivlnfnphfelekhdkydf 87
QY 146 LUF-----TSSSDQVGPYCGSMTVPEKELLMTSEVTFRESGSHISGRFLTY-AS 196
DB 88 ldlrtdgsesadllgkhcgnl-aptlissgsvlyikfidsdyarqagfalyelftkgs 146
QY 197 SD-----HPDLITC-----LEBASHTKTEYSK----- 220
DB 147 edcsknfetspngtlespypfepkypnhdctftllakprmeillqflltdlehpdlqvag 206
QY 221 -CPACRVDAGDIS--GNMVDGYRDTSLCKAIIHAGIADLEGQISVLQKRGISRY-- 275
DB 207 dckykwdldwglpwhvsgpllgkycgktpsklrstqglstlftldmavakdgfsaryyl 266
QY 276 -----ECILANGVLSRD-GSLSDKRFLFTSNGCSRLSFPDQIRASSWQSVNCSQ 328
DB 267 vhppepfqcnaplmgmesgflane-----qlasastf-----s 300
QY 329 QVHMSPGQARLDDQSPNASGDSNNHPRKMLEIDGKKRTIGRTTG--STQSNFNF 386
DB 301 dgwtpqsrllgddngwcpnldsn---keylgvdrlflmltalatqalsretkgyy 356
QY 387 YVKSFYVNFKNNNSKMTYKGIYVNEEFQGNFNFRPQNNFPIVARYVAVVQTHQIALKVE 446
DB 357 yvksykrleavstngedkmmyrgrkh--kifamndatevlnklmpdltrfltrpqlwhglalrle 414
QY 447 HORIALKVELIGCOITGNDSLVWRKTS--QSTSVTKKEDETITRP 491
DB 415 hlglalrltelfgcvtdapcsnmlgmisgladtqiasastreyllwsp 462

RESULT 10
AAB64627
ID AAB64627 standard; Protein; 439 AA.
AC AAB64627;
XX
XX 22-MAR-2001 (first entry)
XX DE Human secreted protein BLAST search protein SEQ ID NO: 137.
XX

Dd	474	ctagccgacgagtgcaccatccaggatttaaacctgttttggaagcagcgcccatattt	533
Qy	639	gaagacgaatacacgcaaatctctgccagctgttgttagagaagtagagagacaatttc	698
Dd	534	cgaagataaaatacagccaattctctccagcctgtgttagagacatatgacgagataatttc	593
Qy	659	tgggaatatgttatatgatataagaataacctcttattttgtcacaagctgcacatcgtc	758
Dd	554	tgggaatacaaaaagatgtgttacagagataacctctttatttctgcacaagctgcacaccagc	653
Qy	759	aggaataatctgtatgaactaggctggccagatcagtgcttcacagcacaagggatcag	818
Dd	654	agggaatctccacagatgacactaggctggccacaatacaactgtcttcagaagcacaaggataag	713
Qy	819	tgcataatgaagaagatctctgcgcaatgtgtctcttccttcgagagatggtctccctgcacaa	878
Dd	714	tcacatgtgaaggaactctcttcgcgcaatgtgcgtctctcccgcacatggtctcttgcgaaaa	773
Qy	879	ggcattctctgtttacctccaatg	902
Dd	774	ggcattctctttacaacccag	797
<hr/>			
RESULT 6			
AAH34855			
ID	AAH34855	standard; cDNA; 2428 BP.	
XX	AAH34855;		
XX	08-SEP-2001	(first entry)	
DT	Hutan cplon cancer antigen encoding cDNA SEQ ID NO:1937.		
DE	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; ss.		
XX	Homo sapiens.		
XS			

	RESULT	6
AAH34855		
ID	AAH34855 standard; cDNA; 2428 BP.	
XX		
AC	AAH34855;	
XX		
DT	08-SEP-2001 (first entry)	
XX		
DE	Human cplon cancer antigen encoding cDNA SEQ ID NO:1937.	
XX		
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection.	
KM	colorectal carcinoma; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	M0200128920-AZ.	
XX		
PD	05-APR-2001.	
XX		

XX	29-SEP-1999.	9905-0157137.
PR	03-NOV-1999.	9905-0163280.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Ruben SM, Barash SC, Blrse CE, Rosen CA;	
XX		
XX	WPI; 2001-235357/24.	
DR	P-PSDB; AAG75450.	
XX		
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,	
PT	useful for preventing, diagnosing and/or treating colorectal cancers -	
XX		
PS	Claim 1; Page 3444-3445; 9803pp; English.	
XX		
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon	
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where	
CC	the proteins are collectively known as colon cancer antigens. The colon	
CC	cancer antigens have cytostatic activity and can be used in gene	
CC	therapy and vaccine production. N and P may be used in the prevention,	
CC	diagnosis and treatment of diseases associated with inappropriate P	
CC	expression. For example, N and P may be used to treat disorders	
CC	associated with decreased expression by rectifying mutations or deletions	
CC	in a patient's genome that affect the activity of P by expressing	
CC	inactive proteins or to supplement the patients own production of P.	
CC	Additionally, N may be used to produce the colon cancer-associated Ps,	
CC	by inserting the nucleic acids into a host cell and culturing the cell	
CC	to express the proteins. N and P can be used in the prevention, diagnosis	
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204	
CC		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 01:13:23 ; Search time 69.09 Seconds

(without alignments)
815.000 Million cell updates/sec

Title: US-09-691-344A-4

Perfect score: 3052

Sequence: 1 MGFGAGQRLRPVAPRASSAF.....YDNEKEMTKLDTSDMAC 586

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	331	10.8	2211	1 KFB05	coagulation factor
2	330	10.8	927	1 J00948	A5 antigen precurs
3	321	10.5	2324	1 KKHU5	coagulation factor
4	318	10.4	2183	2 T42764	coagulation factor
5	299	9.8	218	2 A47285	milk fat globule p
6	297.5	9.7	427	2 JC4915	ags protein precu
7	285.5	9.4	216	2 A44258	factor VIII-assoc
8	285.5	9.4	2351	1 EZHU	coagulation factor
9	276.5	9.1	463	1 A36479	milk fat globule m
10	276.5	9.1	2319	2 A47004	coagulation factor
11	271.5	8.9	2133	2 T42763	coagulation factor
12	268	8.8	401	2 S65138	glycoprotein anti
13	268	8.8	427	2 S74211	PAS-6/7 protein pr
14	259.5	8.5	409	2 T11743	pp47 protein - pig
15	217	7.1	845	2 JC5256	adipocyte transcr
16	216.5	7.1	719	2 S51739	transcription repr
17	192.5	6.3	1283	2 T11799	neurexin IV - frul
18	187	6.1	3133	2 S52093	hemocytin - silkw
19	186.5	6.1	686	1 A59271	Ra-reactive factor
20	174	5.7	1524	2 T30337	polypeptide - Afri
21	172.5	5.7	319	2 T15159	UVS 2 protein - Af
22	168	5.5	3623	2 B09456	intrinsic factor-B
23	165.5	5.4	986	1 B58788	procollagen C-endo
24	165	5.4	730	1 BMHUI	procollagen C-endo
25	165	5.4	823	1 A58788	procollagen C-endo
26	163.5	5.3	449	2 A55362	procollagen I C-pr
27	162.5	5.3	819	2 I48859	tyro 10 receptor k
28	160.5	5.3	694	2 JC6554	complement subcomp
29	158	5.2	991	2 I49540	procollagen C-endo

30	156	5.1	707	2 JC2218	procollagen C-endo
31	156	5.1	3623	2 T08618	intrinsic factor-B
32	155.5	5.1	855	2 S42621	protein-tyrosine k
33	153	5.0	402	2 JH0403	procollagen I C-pr
34	149.5	4.9	695	1 S05008	complement subcomp
35	146.5	4.8	1019	2 A38738	coagulation factor
36	142.5	4.7	1385	2 T14158	neurexin IV - mous
37	141	4.6	688	1 C1HUS	complement subcomp
38	141	4.6	2083	2 T42721	CRP-ductin-alpha p
39	140.5	4.6	1381	2 T31083	paranodin - rat
40	137	4.5	705	1 C1HURB	complement subcomp
41	136.5	4.5	699	1 I54763	Ra-reactive factor
42	135	4.4	737	2 T15615	hypothetical prote
43	133.5	4.4	1594	2 T30549	hensin - rabbit
44	133	4.4	855	2 JC7731	membrane-bound arg
45	133	4.4	855	2 JC7775	membrane type-seri

ALIGNMENTS

RESULT 1

KFB05 coagulation factor V precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999

C:Accession: M42580; A36497

R:Guineto, E.R.; Esmen, G.T.; Mann, K.G.; MacGillivray, R.T.

J. Biol. Chem. 267, 2971-2978, 1992

A:Title: The complete cDNA sequence of bovine coagulation factor V.

A:Reference number: A42580; MUID:92147638

A:Accession: M42580

A:Molecule type: mRNA

A:Residues: 1-2211 <GUI>

A:Cross-references: GB:M81440; NID:g163037; PID:AAA30512.1; PID:g163038

A>Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBIIP:80776)

R:Kafafatis, M.; Jenny, R.J.; Mann, K.G.

J. Biol. Chem. 265, 21580-21589, 1990

A:Title: Identification and characterization of a phospholipid-binding site of bovine

A:Reference number: A36497; MUID:91072354

A:Accession: A36497

A:Molecule type: Protein

A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X'

R:Xue, J.; Kafafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.

Biochemistry 33, 13109-13116, 1994

A:Title: Determination of the disulfide bridges in factor Va heavy chain.

A:Reference number: A55979; MUID:95034740

A:Contents: annotation

A>Note: 566-Cys and 617-Cys were shown to have free sulfhydryls

C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.

C:Function:

A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p

A:Pathway: blood coagulation

C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase

C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plas

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-2211/Product: coagulation factor V #status predicted <MNT>

F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAL>

F:29-345/Domain: A1 <DA1>

F:33-329/Domain: ferroxidase repeat homology <FO1>

F:346-695/Domain: A2 <DA2>

F:351-688/Domain: ferroxidase repeat homology <FO2>

F:696-1564/Domain: B <DOB>

F:1175-1437/Region: 9-residue repeats (O-X-T/N-I-S-P-D-I-S)

F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>

F:1565-1892/Domain: A3 <DA3>

F:1572-1892/Domain: ferroxidase repeat homology <FO3>

F:1654-1752/Region: phospholipid binding #status predicted

F:1893-2051/Domain: C1 <DC1>

F:2052-2211/Domain: discoidin I amino-terminal homology <DN1>

F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>

F:167-193, 248-329, 499-525/Disulfide bonds: #status experimental

E:325-329,297,382,460,553,587,745,756,774,780,902,952,964,1044,1053,1062,1071,1078,1094
F:333-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:362,627,657/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor xa) #status predicted
F:333-334/Cleavage site: Arg-Gly (protein C) #status predicted
F:607-668,1712-1738,1894-2048,2053-2208/Disulfide Bonds: #status predicted
F:741-742/Cleavage site: Arg-Ser (coagulation factor xa, thrombin) #status predicted
F:1034-1035/Cleavage site: Arg-Ser (coagulation factor xa, thrombin) #status predicted
F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match	10.8%;	Score 331;	DB 1;	Length 2211;
Best Local Similarity	32.5%;	Pred. No. 6.1e-16;		
Matches	88;	Conservative	48;	Mismatches 89;
				Indels 46;
				Gaps 10;

QY	209	ASHLAKTEY-SKRCPA-----GCRVADIGSNMY--DGRUDISLCKAAIHAGIADDE	259
Dd	1968	AKHHLKYYTEECVAVSLDRKNMRLEFNSNTRWNYFCGNSDASTIEINOIDPPVA--	2025
QY	260	LGGQISVLQRKGISREGILANGVLSRDSGLDRKFRFLTNSCGRSLFE----PDGOIR	315
Dd	2026	-----RTIRISPQSYNNKPALRLRELOG- EYNCCSITPLDMESGKIENKOIT	2070
QY	316	ASS-----SWGSVNESGDQVHMSPEQAFLDODG--PSWASGDDSNHHKPREMLEIDLGEKK*	370
Dd	2071	ASSSKKSMWG-----NWMEPLATLNAGRYVMQAOKANNENN-----OMLOIDELKLTKK	2119
QY	371	ITGIPTGTOSTQNFNFYSVFNVFNKNSKKTATKGAIVNNEKEFYOGSNSPRDPVOINF	430
Dd	2120	ITALYVGCGCSLSSEMTVSYTHIHSDQGTDMKPIFREKSSMYDKIFEGBNNGVGHVKNFF	2179
QY	431	IPIPVARYRVVPQVWHORIALKVELLICOI	461
Dd	2180	NPPIISKFIPIRIKTYMNOISIAFLLELFEGCD	2210

RESULT 2

A5 antigen precursor - African clawed frog
C/Species: *Xenopus laevis* (African clawed frog)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: JH0466; J00948
R/Takagi, S.; Hiraiz, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
Neuron 7, 295-307, 1991
A>Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homology
A/Reference number: JH0466; MUID:51337458
A/Accession: JH0466
A/Molecule type: mRNA
A/Residues: 1-927 <TM>
A/Cross-references: GB:D10467; GB:D01077; NID:q222962; PIDN:BA01260.1; PID:q222963
A/Experimental source: tadpole, brain
A/Note: this protein has motifs homologous to complement components C1r and C1s and to C
C/Comment: This protein is a neuronal cell surface molecule involved in the neuronal rec
C/Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-terminal h
C/Keywords: duplication; glycoprotein; transmembrane protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-927/Product: A5 antigen #status predicted <A5>
F:27-138/Domain: C1r/C1s repeat homology <C1R1>
F:147-262/Domain: C1r/C1s repeat homology <C1R2>
F:274-424/Domain: discoidin I amino-terminal homology <DN1>
F:430-584/Domain: discoidin I amino-terminal homology <DN2>
F:646-812/Domain: MAM homology <MAM>
F:861-883/Domain: transmembrane #status predicted <TM>
F:150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.8%; Score 330; DR 1; Length 927;
Best Local Similarity 24.28%; Pred. No. 2e-16;
Matches 112; Conservative 74; Mismatches 171; Indels 106; Gaps 20;

QY 66 LIALLLAVSAPRLQAEEILGDCGHLYVYOSGTWTSKNYPTGYNHHVCETKITVPKG- 124
 | | | | | | | | | | | | | | | |
db 5 LLSCCWMILCSIRSSMASRNKDCGPIITPSYSYTLASGPISYPSPORCEILLADPEHY 64
 | | | | | | | | | | | | | | | |

```

0Y 125 KRLILRLG-DLIDIESQTCASLYLF---TSSDOYGPYCGSMTPVKELLNLTSEVTRF 179Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 ORIMNFMPHFDLEDRCKYDVEVIDGDANNGOLLGKYGCKI-APPLVSTGSPIFRF 122X
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 180 ESGHIIISGRGLFLLY-----ASSD-----HPDLITC----- 209Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 VSDYETTPAGSIRIEYFKTPGECSSRNFTSSNGYIKSPKPEKIPNMBDECIYIIFAPKMQ 188X
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 206 ---LEASHYKTEYSKCPAG--CR-DVADIDSGNWDGYRDTSLCKAAIHAGIAD 259Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 ELYLEFSEFELEADBN-APGQYCRDMDLG-----IMDGRPG-----YGPH 222X
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 260 LGGOISYLORKGISRYESILA-----NGVLSRDG-----SLSRKPLEFTSNGCSR 305Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 IGRYCGQMTPGRVNSFTGLTSMIFHTSALIKEGFANFYSVOISNTDEDF-----QCKEA 278Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 306 LSFEPD-----GOIRASSSMQSVNMSGDQVHNSPGQARLPDQCPGSMAGDSSNNHKPREWL 361Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 LGMESGELHPDOISVSOY-----SMNNSAESRRLNYENCMTGCEDT---XKEMI 326Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 362 EIDLDEKKRIKTIIRTYG--STOSNPFNFYKFSVANKNNKSNKMTYKGIYNNNEKVEQGN 419Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 QYDLENLAFYVGIGIGOGALISKEFKKTKYVSKYKVDISNMGEDWTLTD--GNKRLVPTGN 384X
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 420 SNFRDPVONNFIPIVARYVNVVQOTWHQRIALKVELIGCOIT 462Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 TDAATDVVYRPFESKPYITFRVFLRLPVYTWNGSLSEFPLYGCYKIT 427Y
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 3

coagulation factor V precursor [validated] - human

N:Alternate names: coagulation labile factor; proaccelerin

C:Species: Homo sapiens (man)

C:Date: 19-May-1989 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000

C:Accession: A56172; A42344; A28028; A27498; A25897

R:Clp, L.D.; Moore, K.D.; Kane, W.H.

Biochemistry 31, 3777-3785, 1992

A:Title: Structure of the gene for human coagulation factor V.

A:Reference number: A42344; MUID:92232668

A:Accession: A56172

A:Molecule type: DNA

A:Residues: 1-2224 <CR1>

A:Cross-references: GB:J05368

A:Accession: A42344

A:Molecule type: DNA

A:Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-52070;2111-2120;2172-2181 <CR2>

R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Ka

Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987

A:Title: Complete cDNA and derived amino acid sequence of human factor V.

A:Reference number: A28028; MUID:87260886

A:Accession: A28028

A:Molecule type: mRNA

A:Residues: 1-857, 'R',859-864, 'R',866-924, 'E',926-1763, 'I',1165-2212, 'T',2214-2224 <J

A:Cross-references: GB:M16967

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davle, E.W.

Biochemistry 26, 6508-6514, 1987

A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of

A:Reference number: A27498; MUID:88107560

A:Accession: A27498

A:Molecule type: mRNA

A:Residues: 1-1284, 'I',1286-1600 <KAN>

A:Cross-references: GB:M17785

A:Note: parts of this sequence were determined by protein sequencing

R:Kane, W.H.; Davle, E.W.

Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986

A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homo

A:Reference number: A25897; MUID:86313665

A:Accession: A25897

A:Molecule type: mRNA

A:Residues: 1188-1215,1315-2224 <KA2>
A:Cross-references: GB:M14335
A>Note: parts of this sequence were determined by protein sequencing
R.Keller, F.G., Ortel, T.L., Quinn-AlLEN, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; MUID:95210278
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:1F5
A:Cross-references: GDB:119896; OMIM:227400
A:Map position: 1q23-1q23
A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 655/1
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa proteolytic cleavage of fibrinogen to fibrin.
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase reductase; blood coagulation; duplication; glycoprotein; phospholipid binding; plasma; F1-18/Domains: signal sequence #status predicted <SIG>
F:29-2224/Product: coagulation factor V #status predicted <MAT>
F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
F:29-345/Domains: A1 <DA1>
F:33-328/Domains: ferroxidase repeat homology <FO1>
F:346-691/Domains: A2 <DA2>
F:351-684/Domains: ferroxidase repeat homology <FO2>
F:692-1573/Domains: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domains: A3 <DA3>
F:1581-1905/Domains: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domains: C1 <DC1>
F:1906-2064/Domains: discoidin I amino-terminal homology <DN1>
F:2065-2224/Domains: C2 <DC2>
F:2065-2221/Domains: discoidin I amino-terminal homology <DN2>
F:51-55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,1494,1507,1511,1512,1513,1514,1515,1516,1517,1518,1519,1520,1521,1522,1523,1524,1525,1526,1527,1528,1529,1530,1531,1532,1533,1534,1535,1536,1537,1538,1539,1540,1541,1542,1543,1544,1545,1546,1547,1548,1549,1550,1551,1552,1553,1554,1555,1556,1557,1558,1559,1560,1561,1562,1563,1564,1565,1566,1567,1568,1569,1570,1571,1572,1573,1574,1575,1576,1577,1578,1579,1580,1581,1582,1583,1584,1585,1586,1587,1588,1589,1590,1591,1592,1593,1594,1595,1596,1597,1598,1599,1600,1601,1602,1603,1604,1605,1606,1607,1608,1609,1610,1611,1612,1613,1614,1615,1616,1617,1618,1619,1620,1621,1622,1623,1624,1625,1626,1627,1628,1629,1630,1631,1632,1633,1634,1635,1636,1637,1638,1639,1640,1641,1642,1643,1644,1645,1646,1647,1648,1649,1650,1651,1652,1653,1654,1655,1656,1657,1658,1659,1660,1661,1662,1663,1664,1665,1666,1667,1668,1669,1670,1671,1672,1673,1674,1675,1676,1677,1678,1679,1680,1681,1682,1683,1684,1685,1686,1687,1688,1689,1690,1691,1692,1693,1694,1695,1696,1697,1698,1699,1700,1701,1702,1703,1704,1705,1706,1707,1708,1709,1710,1711,1712,1713,1714,1715,1716,1717,1718,1719,1720,1721,1722,1723,1724,1725,1726,1727,1728,1729,1730,1731,1732,1733,1734,1735,1736,1737,1738,1739,1740,1741,1742,1743,1744,1745,1746,1747,1748,1749,1750,1751,1752,1753,1754,1755,1756,1757,1758,1759,1760,1761,1762,1763,1764,1765,1766,1767,1768,1769,1770,1771,1772,1773,1774,1775,1776,1777,1778,1779,1780,1781,1782,1783,1784,1785,1786,1787,1788,1789,1790,1791,1792,1793,1794,1795,1796,1797,1798,1799,1800,1801,1802,1803,1804,1805,1806,1807,1808,1809,1810,1811,1812,1813,1814,1815,1816,1817,1818,1819,1820,1821,1822,1823,1824,1825,1826,1827,1828,1829,1830,1831,1832,1833,1834,1835,1836,1837,1838,1839,1840,1841,1842,1843,1844,1845,1846,1847,1848,1849,1850,1851,1852,1853,1854,1855,1856,1857,1858,1859,1860,1861,1862,1863,1864,1865,1866,1867,1868,1869,1870,1871,1872,1873,1874,1875,1876,1877,1878,1879,1880,1881,1882,1883,1884,1885,1886,1887,1888,1889,1890,1891,1892,1893,1894,1895,1896,1897,1898,1899,1900,1901,1902,1903,1904,1905,1906,1907,1908,1909,1910,1911,1912,1913,1914,1915,1916,1917,1918,1919,1920,1921,1922,1923,1924,1925,1926,1927,1928,1929,1930,1931,1932,1933,1934,1935,1936,1937,1938,1939,1940,1941,1942,1943,1944,1945,1946,1947,1948,1949,1950,1951,1952,1953,1954,1955,1956,1957,1958,1959,1960,1961,1962,1963,1964,1965,1966,1967,1968,1969,1970,1971,1972,1973,1974,1975,1976,1977,1978,1979,1980,1981,1982,1983,1984,1985,1986,1987,1988,1989,1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021,2022,2023,2024,2025,2026,2027,2028,2029,2030,2031,2032,2033,2034,2035,2036,2037,2038,2039,2040,2041,2042,2043,2044,2045,2046,2047,2048,2049,2050,2051,2052,2053,2054,2055,2056,2057,2058,2059,2060,2061,2062,2063,2064,2065,2066,2067,2068,2069,2070,2071,2072,2073,2074,2075,2076,2077,2078,2079,2080,2081,2082,2083,2084,2085,2086,2087,2088,2089,2090,2091,2092,2093,2094,2095,2096,2097,2098,2099,2100,2101,2102,2103,2104,2105,2106,2107,2108,2109,2110,2111,2112,2113,2114,2115,2116,2117,2118,2119,2120,2121,2122,2123,2124,2125,2126,2127,2128,212

```

A:Title: The structure and function of murine factor V and its inactivation by prote
A:Reference number: Z22270; MUID:98282202
A:Accession: T42764
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-2183 <YAN>
A:Cross-references: EMBL:U52925; NID:g3219690; PID:g3219691; PIDN:AAC9553.1
C:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plas
F:350-682/Domain: ferroxidase repeat homology <FOX1>
F:1541-1864/Domain: ferroxidase repeat homology <FOX2>

Query Match 10.4%; Score 318; DB 2; Length 2183;
Best Local Similarity 40.8%; Pred. No. 5.7e-15;
Matches 69; Conservative 31; Mismatches 53; Indels 16; Gaps 5;

OY 300 NGCSRSLSFE---PDGQIRASS---SMQSVNMGDDQVHMSPGQARLQDQSPSWASGDS 352
||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 2023 NGCSPPLGLEDRIDQDKQITFASPEKKSWM-----GD--YMEPSLARLMAQGRVANMAQAKA 2075
||||| : : : : : ||||| : : : : : ||||| : : : : :

OY 353 NNHRRREMLEIDLGKKKITGTGRTGTSQSNFNRYKVSFVNMFKNNSKMTTYGIVNNE 412
||| : : : : : ||| : : : : : ||||| : : : : : || : : : : :
DB 2076 NNNK--QWLQVLDLKKIKVTALVTOGCKSLSEMYKYVSIOYSDQGVAMKPRYKOSMV 2133
||| : : : : : ||| : : : : : ||||| : : : : : || : : : : :

OY 413 EKVEFGNSNFRDPVQNNFIPPIVARYRVVPQTMHORIALKVELIGCOI 461
||| : : : : : ||| : : : : : ||||| : : : : : ||| : : : : :
DB 2134 DKIFEGNSNTKGHMKNFNPPIISRFIRIIRKTNQSIALRLFLFGDI 2182
||| : : : : : ||| : : : : : ||||| : : : : : ||| : : : : :

RESULT 5
A47285 milk fat globule protein - human (fragment)
C:Species: Homo sapiens (man)
C:Dates: 31-Dec-1993 #sequence_rev1sion 31-Dec-1993 #text_change 28-May-1999
C:Accession: A47285
R:Ratocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.I.
Cancer Res. 51, 4994-4999, 1991
A:Title: A M-r 46,000 human milk fat globule protein that is highly expressed in huma
A:Reference number: A47285; MUID:91371351
A:Accession: A47285
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <LAR>
A:Cross-references: GB:S56151; NID:g2325396; PIDN:AAB19771.1; PID:g2325397
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F:60-218/Domain: discoidin I amino-terminal homology <DN2>

Query Match 9.8%; Score 299; DB 2; Length 218;
Best Local Similarity 37.3%; Pred. No. 5e-15;
Matches 63; Conservative 33; Mismatches 55; Indels 18; Gaps 5;

OY 300 NGCSRSLSFE---PDGQIRASSSMQSVNMGDDQVH---MSPGQARLQDQSPSWASGD 350
||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 59 NGCANPLPGLKNNISIPDKQITFASSSKYTWG-----LHLEFNMPSYARLDQGNFAMVAGS 113
||||| : : : : : ||||| : : : : : ||||| : : : : :

OY 351 SSNNHRRREMLEIDLGKKKITGTGRTGTSQSNFNRYKVSFVNMFKNNSKMTTYKGIYN 410
||| : : : : : ||| : : : : : ||||| : : : : : || : : : : :
DB 114 YGND---QWLQVLDLGGSSKEYVTGIITOGARNFQSVQFVASYKVAYSNDASAMWTEYQDPRT 169
||||| : : : : : ||| : : : : : ||||| : : : : : ||| : : : : :

OY 411 NEEKVEFGNSNFRDPVQNNFIPPIVARYRVVPQTMHORIALKVELIGC 459
||||| : : : : : ||| : : : : : ||||| : : : : : ||| : : : : :
DB 170 GSSKIFPGNMNHSKKNLEFETPIIARVRLPAMHNRIALRLLELGC 218
||||| : : : : : ||| : : : : : ||||| : : : : : ||| : : : : :

RESULT 6
JG4915 aggs protein precursor - rat
N:Alternate names: O-acetyl-Gd3 ganglioside

```

F:57-210/Domain: discoidin I amino-terminal homology <DN2>

A:Title: Identification and functional importance of tyrosine sulfate residues within
A:Reference number: A42348; MUID:92207952
A:Accession: A42348

A:Molecule type: protein
A:Residues: 20-36;356-371;392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709-1710
A:Experimental source: recombinant material from Chinese hamster ovary cells
A>Note: Sequence extracted from NCBI backbone and corrected to correspond with the published sequence.
R:Fay, P.J.; Smudgin, T.M.
J.Biol.Chem. 264, 14005-14010, 1989
A>Title: Inter-subunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500
A:Accession: A43986

A:Molecule type: protein
A:Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Hutter, W.B.; Verbeeck, M.P.; Mertens, K.
J.Biol.Chem. 266, 740-746, 1991
A>Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for its biological activity.
A:Contents: annotation; sulfation
A:References: annotation; sulfation
R:Gileschier, J.; Wood, W.T.; Goralica, T.M.; Wilson, K.L.; Chen, E.Y.; Eaton, D.H.; Vohar, R.
Nature 312, 326-330, 1984
A>Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547
A:Contents: annotation; introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A>Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of human factor VIII.
A:Reference number: A56216; MUID:95338127
A:Contents: annotation; disulfide bonds
A>Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
R:Kjalke, M.; Hedner, U.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
Eur.J.Biochem. 234, 773-779, 1995
A>Title: Amino acid residues 721-729 are required for full factor VIII activity.
A:Reference number: S63527; MUID:96163459
A:Accession: S63527

A:Molecule type: protein
A:Residues: 733-752;753-759 <KJA>
R:Rind, P.; Larsson, K.; Spyra, J.; Sydow-Baeckman, M.; Almsstedt, A.; Gray, E.; Sandberg, E.
Eur.J.Biochem. 232, 19-27, 1995
A>Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction and characterization.
A:Reference number: S66445; MUID:96046024
A:Accession: S66445

A>Status: Preliminary
A:Molecule type: protein
A:Residues: 1668-1685 <LNT>
C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure produces inactive fragments.
C:Genetics:

A:Gene: GDB:FBC
A:CROSS-references: GDB:119124; OMIM:306700
A:Map position: Xq28-Xq28
F:1>Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 633/1
C:Function:

A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa proteolytic cleavage of prothrombin to thrombin.
A:Pathway: blood coagulation
C:Superfamily: coagulation factor VII; discoidin I amino-terminal homology; ferroxidase C keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; platelet activation; signal sequence #status predicted <SIG>
F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>
F:23-356/Domains: A1 <DAI>
F:23-348/Domains: A1 <DAI>
F:23-348/Domains: ferroxidase repeat homology <FOI>
F:392-759/Domains: A2 <DAA>
F:402-730/Domains: ferroxidase repeat homology <FO2>
F:760-1667/Domains: B <DBO>
F:1668-3351/Product: coagulation factor VIIIA light chain #status experimental <ACLI>
F:1709-2038/Domains: A3 <DAB>
F:1716-2038/Domains: ferroxidase repeat homology <FO3>
F:2039-2191/Domains: C1 <DC1>
F:2039-2188/Domains: discoidin I amino-terminal homology <DN1>
F:2192-3351/Domains: C2 <DC2>
F:2192-3345/Domains: discoidin I amino-terminal homology <DN2>
F:602-256,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1417,1472,1488,1497,1548,1549,1573,1649,1673,1681,1687,1691,1692,1694,1698,1704,1705,1706,1707,1708,1709,1710,1711,1712,1713,1714,1715,1716,1717,1718,1719,1720,1721,1722,1723,1724,1725,1726,1727,1728,1729,1730,1731,1732,1733,1734,1735,1736,1737,1738,1739,1740,1741,1742,1743,1744,1745,1746,1747,1748,1749,1750,1751,1752,1753,1754,1755,1756,1757,1758,1759,1760,1761,1762,1763,1764,1765,1766,1767,1768,1769,1770,1771,1772,1773,1774,1775,1776,1777,1778,1779,1780,1781,1782,1783,1784,1785,1786,1787,1788,1789,1790,1791,1792,1793,1794,1795,1796,1797,1798,1799,1800,1801,1802,1803,1804,1805,1806,1807,1808,1809,1810,1811,1812,1813,1814,1815,1816,1817,1818,1819,1820,1821,1822,1823,1824,1825,1826,1827,1828,1829,1830,1831,1832,1833,1834,1835,1836,1837,1838,1839,1840,1841,1842,1843,1844,1845,1846,1847,1848,1849,1850,1851,1852,1853,1854,1855,1856,1857,1858,1859,1860,1861,1862,1863,1864,1865,1866,1867,1868,1869,1870,1871,1872,1873,1874,1875,1876,1877,1878,1879,1880,1881,1882,1883,1884,1885,1886,1887,1888,1889,1890,1891,1892,1893,1894,1895,1896,1897,1898,1899,1900,1901,1902,1903,1904,1905,1906,1907,1908,1909,1910,1911,1912,1913,1914,1915,1916,1917,1918,1919,1920,1921,1922,1923,1924,1925,1926,1927,1928,1929,1930,1931,1932,1933,1934,1935,1936,1937,1938,1939,1940,1941,1942,1943,1944,1945,1946,1947,1948,1949,1950,1951,1952,1953,1954,1955,1956,1957,1958,1959,1960,1961,1962,1963,1964,1965,1966,1967,1968,1969,1970,1971,1972,1973,1974,1975,1976,1977,1978,1979,1980,1981,1982,1983,1984,1985,1986,1987,1988,1989,1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021,2022,2023,2024,2025,2026,2027,2028,2029,2030,2031,2032,2033,2034,2035,2036,2037,2038,2039,2040,2041,2042,2043,2044,2045,2046,2047,2048,2049,2050,2051,2052,2053,2054,2055,2056,2057,2058,2059,2060,2061,2062,2063,2064,2065,2066,2067,2068,2069,2070,2071,2072,2073,2074,2075,2076,2077,2078,2079,2080,2081,2082,2083,2084,2085,2086,2087,2088,2089,2090,2091,2092,2093,2094,2095,2096,2097,2098,2099,2100,2101,2102,2103,2104,2105,2106,2107,2108,2109,2110,2111,2112,2113,2114,2115,2116,2117,2118,2119,2120,2121,2122,2123,2124,2125,2126,2127,2128,2129,2130,2131,2132,2133,2134,2135,2136,2137,2138,2139,2140,2141,2142,2143,

F:414_426/Binding site: sulfate (Tyr) #status predicted
F:755-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1657-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1740-1741/Cleavage site: Arg-Ser (coagulation factor Xa) #status experimental
F:2193-2345/Disulfide bonds: #status predicted

	Query Match	9.48	Score 285.5	DB1	Length 2351
	Best Local Similarity	37.68	Pred. No. 17e-12		
	Matches 62	Conservative 32	Mismatches 58	Indels 13	Gaps
QY	300	NGCSRSLSE----	PDQGIASSSMQSVNESCQVWVHSPGARLDQCGSPWAGDSSNNH	355	
Db	2191	NSCSMPPLGMEKALSDQIATASSYFTNMFAT-----	WSSKARLHLOGSNMNRQVYN-	2244	
QY	356	KPRLMEIDIGKKKITGITRTGSGTOSNFNFYKSFVNMFKNNSKWKTYGIVNEEKV	415		
Db	2245	-PKEMLDVDFQKTKVYGVTTQGVKSLTSMYKKEPLISSDDGHQWTLF--	FQNGKVKV	2301	
QY	416	FQGSNSRDPYNNFPIPIYARVYRVYPCQVHRIALKEVLIGCQ	460		
Db	2302	FQGGQDSFTPVVNSLDPEPLLTTRLRLRHPOSWAQIILREVLIGE	2346		

RESULT 9
A36479
milk fat globule membrane protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A36479
R:Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bul, A.; Yuzuki, D.; Sriniwasan, U.; Parry
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the
A:Reference number: A36479; MUID:91046008
A:Accession: A36479
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-463 <STU>
A:Cross-references: GB:M38337; NID:g199142; PIDN:AAA39534.1; PID:g199143
A:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C:Keywords: membrane protein
E:28-60/Domain: EGF homology <EG1>
E:68-107/Domain: EGF homology <EG2>
E:147-303/Domain: discoidin I amino-terminal homology <DN1>
E:307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match	9.1%	Score 276.5	DB 1	Length 463
Best Local Similarity	23.4%	Fred. No. 7.5e-13		
Matches 112	Conservative 59	Mismatches 118	Indels 189	Gaps 21
QY	141	CASDYLFTSSDDXGAPCGSMWYVKELLINTSEVTVPRESGSHISGR-----GF	190	
Db	16	CASG--LPAASD-----FCDS-----SLCLN-----GGTCLTGQDNDIYCLCPGCF	55	
QY	191	LLTYAS-----SDHPDLITCLERASHYLKTEYSKFCPAGCRDVAIGDISGN-	236	
Db	56	TGLVNETETERGCPSENPQYDAKCLVT-LDPTGRGIDLPFEYIQGCPVYSGINCHEETNY	114	
QY	237	VDG-YRDTSLICKRAI-----HAGITAD-----	258	
Db	115	NLDGSEYMTTAVPNPNAVPTPAFTPDLISNNLASRCSSTOLGMEGAGLADISASYVMGM	174	
QY	259	---ELGGISYLORGI-----SREGILANGVLSROG-----	288	
Db	175	GLQRMGPPLARLYRTGILYNNAMHASWYDSKPMYIQVLLKRMKRVSGWTOGA-SRAGRAEYL	233	
QY	289	---SLSDKRFLFTSN-----GCSR-304		
Db	234	KTFKVAAYLSDGRKPEFIDGSGDKEFLGINDNNSLTKYMMFNPTLEAOYIRILPYPSCHRG	293	
QY	305	---SLSFE-----PDGQIRASSWSQVNESGQVHMSPGQARLADOG	343	

```

D0 CTLEFELLGCGELHGLCEPLGLEKLNNTIPDSQASSSSYKTWMLRA--FGMYPHLGRLDNQG 351
      :|||
D1 P- -SWASGDSSNNHKPRFWLIDIGEEKKINGIFTGTSGNSFNPNVYSPFMNFENNNKK 401
      :|||
D2 KINAW----TKQSSAKMQLDYLGSTORQVTGIITQGRDFGHDIQYVSYSVAHSDDGVQ 407
      :|||
OY 402 WKTYKGIVNNEKEVFQGSNFRDPVNQNNFIPIVARVVVPQTWHORIATLKVELIGC 459
      :|||
Db 408 WTVYE--EQGSSKFQOGLDNNSHKKNFEKPFMARVRYLPVSMHNRIITRELLGC 463
      :|||

RESULT 10
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
C:Accession: A47004
R:Eider, B.; Laskich, D.; Gletscher, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII cDNA.
A:Reference number: A47004; MUID:93300511
A:Accession: A47004
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <GLD>
A:Cross-references: GB:I05573; NID:g192456; PIDB:AAA37385.1; PID:g192457
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:23-349/Domain: ferroxidase repeat homology <FO1>
F:402-730/Domain: ferroxidase repeat homology <FO2>
F:1886-2006/Domain: ferroxidase repeat homology <FO3>
F:2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>
```

```

Query Match          9.1% Score 276.5; ID 2; Length 2319;
Best Local Similarity 86.3%; Pred. No. 12;
Matches 61; Conservative 33; Mismatches 61; Indels 13; Gaps 4;

QY   300 NGCSRSLSE---PDGQIRASSSMQSYNESGDVHMSPGCARLDQGPMSAGDSSNNH 355
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    2159 NSCSIPLOMESKVIISDPQITASSYFTNMFAT-----NSPSGARLHLGGRTNAMRPQVND-.2212

QY   356 KPRMLELDLGKKIKGICIRTGSTQSNEFNFYASPFVNENNNSKAKTYGIYNNEKV 415
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    2213 -PKQMLVDLTOKTKRVGTILIQGVKSLETFSMFEVEFLISSQDDHHWQI--LVNGKVKV 2269

QY   416 FQGSMNFDPPVQNNEPIPIVARVRYVVQPTMHORIALKVELIGQITQ 463
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    2270 FQGMDSTPMANSLDPPLLRTRYLRHPQIWEOHIALKLLETIGEAOO 2317

RESULT 11
T42763
coagulation factor VIII precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42763
R:Icollar, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z22269
A:Accession: T42763
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2133 <LOC>
A:Cross-references: EMBL:U99517; NID:g1511633; PID:g1511634; PIN:AAB06705.1
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; platelet
F:1-19/Domin: signal sequence <status predicted >SIG>
F:20-2133/Product: coagulation factor VIII <status predicted >MNT>
F:23-349/Domin: ferroxidase repeat homology <FOX1>
F:402-730/Domin: ferroxidase repeat homology <FOX2>

```

```

F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match      8.9%: Score 271.5; DB 2; Length 2133;
Best Local Similarity 35.9%: Pred. No. 1,7e-11;
Matches 61; Conservative 29; Mismatches 57; Indels 23; Gaps 5;

QY 300 NGCSRSLSFE---PDGOIRASSSMQSYNESGDOVHMSFGAQLQDQSPWASGDDSSNNH 355
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1973 NSCSMPLOMOKRAISDQITASSHLSNIFAF-----SPSGARHLQG-----RNNAW 2020
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 356 KPR-----EWLEIDLGEKKKITTGRTGSGTQSNFNFYKFSVMMFKNNNSKWKYKGINV 410
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2021 RPRVSSAEGMLQVDLQKTVKLTGTTCGVKSLASMYVKKEFLVSSQDGRRWTLF--LQD 2078
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 411 NEEVFOGNSNFRDPVONNFTPIPVARVRVRVPCPTMQRILKIKELIGCO 460
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2079 GHTKVFQGNQDSSTPVYNALDPLFTFKRLRHPSMAOHILRLREVLQCE 2128
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 12
S65138
glycoprotein antigen MCP57/53, mammary gland - bovine (fragment)
N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E
C:Species: Bos primigenius taurus (cattle)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C:Accession: S65138; #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A:Title: Molecular cloning of glycoprotein antigens MCP57/53 recognized by monoclonal
A:Reference number: S65138; MUID:96125736
A:Accession: S65138
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-401 <AOK>
R:Mathier, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
II-like sequences.
A:Reference number: A48394; MUID:93250576
A:Accession: G48394
A:Status: preliminary
A:Molecule type: protein
A:Residues: 207-220 <MAT>
A:Experimental source: milk
A:Note: sequence extracted from NCBI backbone (NCBIP:131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C:Keywords: glycoprotein
F:1-32/Domain: EGF homology (fragment) <EG1>
F:40-79/Domain: EGF homology <EG2>
F:82-239/Domain: discoidin I amino-terminal homology <DN1>
F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match      8.88; Score 268; DB 2; Length 401;
Best Local Similarity 36.7%; Pred. No. 2.6e-12;
Matches 61; Conservative 30; Mismatches 63; Indels 12; Gaps 4;

QY 300 NGCSRSLSE----PDGIIRASSSMQSVNESCDDVHNSPGQARRLQDGP--SWASGSSN 353
      |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 242 NGCTEPRLGLKNTIPNKOITRSTSYKTTWGLSA--FSMFPIYARLDNNGKENAWTAQTNSA 299
      |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 354 NHRPREMLEIDLGKKKKTGTGRTGSTQSNPNFYKSFVNMFKNNNSKMKTYKGINNEE 413
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 300 S-----ENLQIDLSGSKKVTGIIITGARDPFGIYQVAAYRAVAGDGYTWTFYKPGASFS 355
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 414 KVEGQNSNFRDPVONNFIPIVARYVRVDPQTWHQIALKVELLGC 459
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 356 KIFPGNMNDNNSHKKNIFETPFOARFVRIOPIVAMHNRIITLRLVLLGC 401

RESULT 13
S74211

```

PAS-6/7 protein precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
A:Accession: S74211; S78114; S24181
R:Havreggaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A>Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat globule
A:Reference number: S74211; MUID:97008954
A:Accession: S74211

A:Molecule type: mRNA
A:Residues: 1-427 <HY>
A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779
A:Accession: S78114

A:Molecule type: protein
A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
R:Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A>Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
A:Reference number: S23926; MUID:9235107
A:Accession: S24181

A:Molecule type: protein
A:Residues: 383-394 <KM>
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolog
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F:1-18/Domains: signal sequence #status predicted <SIG>
F:19-427/Product: PAS-6/7 protein #status experimental <MAT>
F:24-58/Domains: EGF homology <EG1>
F:66-105/Domains: EGF homology <EG2>
F:108-265/Domains: discoidin I amino-terminal homology <DN2>
F:266-427/Domains: discoidin I amino-terminal homology <DN2>
F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted
F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F:59;227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:109-285;252-256;270-427/Disulfide bonds: #status experimental

Query Match 8.8%; Score 268; DB 2; Length 427;
Best Local Similarity 36.7%; Pred. No. 2.9e-12;
Matches 61; Conservative 30; Mismatches 63; Indels 12; Gaps 4;

Oy 300 NGCSRLSE---PDGIRASSQSYNESGDQVHMSPGARLDQGP--SMAGDSSN 353
|||:::|-:-:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 268 NGCTPDLKONTPTPKOTITASSYKTWGLSA--FSNPYYARLDNQEKFNAMTQTNSA 325
|||:::|-:-:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Oy 354 NHKEPMLEIDLGEKKTKTGIRTGSTOSNFNYKSFVNPFKNNSKMKTYKGIVNNEE 413
: |||:::|-:-:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 326 S---EMQLDLDGSKRVGTGIITGARDFGHIQYVAARAVALGDDGVWTETKDPGASES 381
|||:::|-:-:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

Oy 414 KVPQSNFRRDPVNNFTPIPVARYRVVYPQTHORIALKYELIGC 459
|||:::|-:-:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 382 KIFPGNMNDNMSHKKNIFETPFQARFVRIOPAVMHRIILRYELLGC 427
|||:::|-:-:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 14
T11743

PP47 protein - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
A:Accession: T11743
R:Emsslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Troepfeler,
Biol. Reprod. 58, 1057-1064, 1998
A>Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zc
A:Reference number: Z17325; MUID:98206817
A:Accession: T11743

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-409 <ENS>
A:Cross-references: EMBL.Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
A:Experimental source: testis
C:Function:
Description: may be involved in membrane remodeling and/or function as a zona pellucida
C:Superfamily: unassigned EGF-related proteins; EGF homology

```

F:6-40/Domain: EGF homology <EGF>

Query Match          8.5%, Score 259.5; DB 2; Length 409;
Best Local Similarity 25.0%; Pred. No. 1.2e-11;
Matches 87; Conservative 46; Mismatches 102; Indels 113; Gaps 11;

OY 215 TEYSKFCACGCRDVAAGDLSGNVNDGYRDTSL-----LCAAAIHAGII-----ADELGG---- 262
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 72 TEYICKCPHGYTGICHCITCNAPLIGMETGALDFQISASSMHLGFMGIQWAPLRLRHR_131
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 263 -----QISVLQRKGISREYGLIANVLSPDGL----- 290
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 132 AGIVAMTASNTDRMPVQLVNLRR---MRYTGVYTQGA-SRAGSAEYMKTEKVAYSTDG 187
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 291 -----SDKRFL-----FTSNGC 302
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 RKFOPIQGAEESGDKIFMGNDLNSGLKYNLEFVPLEQYVRLVPLICHRGCTLRPELLGC 247
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 303 SRSLSFEEDG-----QIRASSMSQSVNESGQVWMSPGQALDQGP---SWASGDS 351
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 ELISGCAEPLGLKDNTPINKQILTASSFYRLTWGLSA--FSWYPRYALDNGKFNAMTAQSN 305
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 352 SNMHKPREMLEIDLEKKKKTIGRTGSTOSNFNFYKSFVNFEKNNSKMKTYGIVNN 411
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 306 SAS-----EMLOIDLDSQRRVTGITIQGARDPQHIOYVAIKVAYSDDGVSMTFYADQAL 361
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 412 EKVYFQGNNSNFRDPVQNNFTPIYARVYRVVPOTWQHORIAKVELLIGC 459
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 EKGIPGNLNDNNSHKKNMEETPEFLRFVRLPVAAMHNRLTLRVELLGC 409
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
JC5256
adipocyte transcription factor, AEBP1 - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JC5256
R:Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo,
Biochem. Biophys. Res. Commun. 228, 411-414, 1996
A:Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expri
A:Reference number: JC5256; MUID:97079196
A:Accession: JC5256
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-845 <OHN>
A:Cross-references: DDBJ:D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Query Match          7.1%, Score 217; DB 2; Length 845;
Best Local Similarity 25.7%; Pred. No. 5.3e-08;
Matches 55; Conservative 40; Mismatches 97; Indels 22; Gaps 5;

OY 311 DGOIRASSSMO-SVNESGDQVHMSPGQARLDQGSWASGDSNNHKRREMLDLEKK 369
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 84 DNOIRASSMLRGLGAGQGRGLMMQGTATEDDDYDGAACEDART---QWLEVDTRRTT 139
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 370 KITGIRTGSTQSNFNFYKSFVNFEKNNSKMKTYKGIANNNEKRVFQGSNFRDPVQNN 429
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 RTGYITIGGRDSSIHDDVYTTTFYFVGFSDSQTWWTN--GYEEMTFPGNVDKDTPLVSE 197
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 430 FIPPIYARVYRVVPQTMQHORIAKVELIGCOITQ-----GNDSLVWR---KTS 474
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 198 LPEYVAVAFIRIPLTWNGSLCMREVLVGCVAAPYSYAQNVEYATDLDLFRHSHYKDM 257
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
OY 475 QSTSYSTKKDEDTITRPIRPESETSGINTIYAI 508
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
DB 258 ROLMKVNEECPTITRTYSLGKSSRGLKIYAMEI 291
      ||| ||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: July 9, 2002, 02:16:44
Job time: 3801 sec

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 02:15:23 ; Search time 35.54 Seconds

(Without alignments)
638,426 Million cell updates/sec

Title: US-09-691-344A-4

Perfect score: 3052
Sequence: 1 MGFGAGQLRPVPAVRSSAE.....YDNEKEMTKDLITSDWAG 586

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	351.5	11.5	925	1 NRP2_RAT	035276 Rattus norv
2	349	11.4	931	1 NRP2_HUMAN	060462 homo sapien
3	343.5	11.3	931	1 NRP2_MOUSE	035375 mus musculu
4	338	11.1	914	1 NRP1_CHICK	P79795 gallus gall
5	335	11.0	923	1 NRP1_HUMAN	014786 homo sapien
6	331	10.8	922	1 NRP1_RAT	090799 Rattus norv
7	331	10.8	2211	1 FA5_BOVIN	Q28107 bos taurus
8	330	10.8	928	1 NRP1_XENLA	P28824 xenopus lae
9	329	10.8	923	1 NRP1_MOUSE	P97333 mus musculu
10	317	10.4	2224	1 FA5_HUMAN	P12259 homo sapien
11	314	10.3	2258	1 FA5_PIG	094181 sus scrofa
12	299	9.8	427	1 MFGM_HUMAN	008431 homo sapien
13	297.5	9.7	427	1 MFGM_RAT	P70490 Rattus norv
14	285.5	9.4	2351	1 FA8_HUMAN	P00451 homo sapien
15	276.5	9.1	463	1 MFGM_MOUSE	P21956 mus musculu
16	276.5	9.1	2319	1 FA8_MOUSE	006194 mus musculu
17	271.5	8.9	2133	1 FA8_PIG	P12263 sus scrofa
18	268	8.9	427	1 MFGM_BOVIN	095114 bos taurus
19	259.5	8.5	409	1 MFGM_PIG	P79385 sus scrofa
20	229.5	7.5	280	1 XLRI_FUGRU	Q9W675 fugu rubrip
21	221.5	7.3	224	1 XLRI_MOUSE	092114 mus musculu
22	218.5	7.2	224	1 XLRI_HUMAN	015537 homo sapien
23	197	6.5	1308	1 CTA4_HUMAN	090840 homo sapien
24	195.5	6.4	1331	1 CTA4_HUMAN	090840 homo sapien
25	194.5	6.4	1310	1 CTA4_MOUSE	099647 mus musculu
26	192.5	6.3	1284	1 NRX4_DROME	Q94867 drosophila
27	187	6.1	3133	1 HMCT_BOVMO	P98092 bombyx mori
28	186.5	6.1	686	1 MAS2_HUMAN	000187 homo sapien
29	184.5	6.0	1154	1 CTA3_HUMAN	Q9b276 homo sapien
30	172.5	5.7	986	1 UVS2_XENLA	P42664 xenopus lae
31	165	5.4	986	1 BMP1_HUMAN	P13497 homo sapien
32	163	5.3	468	1 PCO1_RAT	008628 Rattus norv
33	162.5	5.3	704	1 CRAR_MOUSE	P98064 mus musculu

34	162.5	5.3	854	1 DDR2_MOUSE	062371 mus musculu
35	161.5	5.3	449	1 PCO1_HUMAN	015113 homo sapien
36	160	5.2	468	1 PCO1_MOUSE	061398 mus musculu
37	159	5.2	550	1 COCH_HUMAN	043405 homo sapien
38	158	5.2	991	1 BMP1_MOUSE	P98063 mus musculu
39	158	5.2	1022	1 TLD_BRARE	057460 brachydanio
40	157	5.1	552	1 COCH_MOUSE	062507 mus musculu
41	156	5.1	707	1 BMP1_XENLA	P98070 xenopus lae
42	155.5	5.1	855	1 DDR2_HUMAN	016832 homo sapien
43	149.5	4.9	695	1 CASP_MESAU	P15156 mesocricetu
44	146.5	4.8	1019	1 LFC_TACTR	P28175 tachyples
45	142.5	4.7	1385	1 CTA1_MOUSE	054991 mus musculu

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	925 AA.
NRP2_RAT				
ID	NRP2_RAT			
AC	035276;			
DT	16-OCT-2001 (rel. 40, Created)			
DT	16-OCT-2001 (rel. 40, Last sequence update)			
DT	16-OCT-2001 (rel. 40, Last annotation update)			
DE	Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).			
GN	NRP2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SPRAGUE-DAWLEY;			
RX	MEDLINE=97433065; PubMed=9288754;			
RA	Kolodkin A.L., Levengood D.V., Rowe E.G., Tai.Y.-T., Giger R.J.,			
RA	Ginty D.D.;			
RT	Neuropilin is a semaphorin III receptor.;			
RL	Cell 90:753-762(1997).			
CC	- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165			
CC	- AND VEGF-145 ISOMERS OF VEGF, AND THE PLGF-2 ISOMER OF PGF.			
CC	- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE			
CC	CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE			
CC	LINING IN THE RIBS.			
CC	- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.			
CC	- SIMILARITY: CONTAINS 2 CUB DOMAINS.			
CC	- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.			
CC	- SIMILARITY: CONTAINS 1 MAM DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: AF016297; AAC53338.1; -			
DR	InterPro: IPR000859; CUB.			
DR	InterPro: IPR000421; FA5B_C.			
DR	InterPro: IPR000998; MAM.			
DR	Pfam: PF00431; CUB; 2.			
DR	Pfam: PF00754; F5_F8_type_C; 2.			
DR	Pfam: PF00629; MAM; 1			
DR	SMART: SM00042; CUB; 2.			
DR	SMART: SM00231; FA5BC; 2.			
DR	SMART: SM00137; MAM; 1.			
DR	PROSITE: PS01180; CUB; 2.			
DR	PROSITE: PS01285; FA5BC_1; 2.			
DR	PROSITE: PS01286; FA5BC_2; 2.			
DR	PROSITE: PS50060; MAM_2; 1.			
DR	Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.			
KW				

```

FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 925 NEUROFILIN-2.
FT DOMAIN 23 858 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 859 883 POTENTIAL.
FT DOMAIN 884 925 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 142 142 CUB 1.
FT DOMAIN 149 267 CUB 2.
FT DOMAIN 277 427 F5/8 TYPE C 1.
FT DOMAIN 434 592 F5/8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT CARBOHYD 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 833 833 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 925 AA; 103896 MM; 3BF62903F644851C CRC64;

```

```

Query Match 11.5%; Score 351.5; DB 1; Length 925;
Best Local Similarity 24.4%; Pred. No. 7.2e-19;
Matches 114; Conservative 76; Mismatches 181; Indels 97; Gaps 17;

```

```

OY 88 CGLHLYTQDSGTMTSKNRYGTYPNHYCEKTLTPK-GKRLILRLG-DLIDISQVCADSY 145
   || :||: ||| ||| ||| ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 28 CGRLMSKAGYITTSFGYDYPHONCEWVYAPENPKYILNPNHEIEKHCKYDF 87
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 146 LFF---TSSDQGYPCYCSMTVPKELLTSEVTVPFESSGSHISGRGLTY-----AS 196
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 IEIRDDSSADLLGHCNI-APPTIISGSVLVIKFTSDARAGAGSLKYLEIPKGS 146
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 197 SD-----HPDLITC---LEBASHTKTEYSK----- 220
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 147 EDCSKNFTSPNCTIPSPGPEKYPHNLDTFTLLAKPMEILLQFLTFLEHDPLOVGBG 206
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 221 -CPACRDVAGDIS--GNVVDGYRDTSLCKAHAGITADELGOISVLQKKGISRY-- 275
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 DCKYMDLWDGIDPHVGPLIGYCGTKPSPKLRSSGTGILSTFPHDMAAKGFSARYYL 266
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 276 -----EGILANGVLSRD-GSLSDKRFLETSNGCSRSLSEFPDQIRASSWQSVNESGD 328
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 VHQEPPEPNCQAPLGMESGRIANE-----QISASSTP-----S 300
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 329 QVHNSPGARLQDQGPSNASGSSNNHKKPREWLELDLCKKKITGIRTTG--STQSNFNF 386
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 DGRWTPQOSRLKGDNGMTPNVDSN---KEYLDVLRFLMTLTAITQGAISRETKQGY 356
   :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 387 YKFSVPMNKNNSKWKTKGIVNNEKVEGNSNFRDPVQGNFIPPIVARYVRVPTW 446
   ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 YKSKLEKLEVTNGEDMMYRRKKNH-KVFOANNDATELVLKLTPLTRIRIRPQW 414
   ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 447 HORIALKVELIGCQITQGNDSLWPKTS---OSTSVSTRKKEDEITTPR 491
   ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 HGLIALRLRELFCGRVTLDAPCSNMLGMLSLADTDQISASSSTREYLMSP 462
   ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 2
NRP2_HUMAN STANDARD: PRT; 931 AA.
ID NRP2_HUMAN
AC 060462; 014820; 014821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
DE NRP2 OR VEGF165R2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
ON [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RA MEDLINE=97470888; PubMed=9331348;
RT Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neurofilin-2, a novel member of the neurofilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RC TISSUE=Breast;
RX MEDLINE=98188099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-O., Neufeld G., Klagesbrunn M.;
RT "Neurofilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neurofilin-2 and neurofilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neurofilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF.";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -1- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A22 (SHOWN HERE), A0 AND A17;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF022859; AAC51788.1; .
DR EMBL; AF022860; AAC51789.1; .
DR EMBL; AF016098; AAC12922.1; .
DR MIM; 602070; .
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_Type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS50060; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20 OR 22 (POTENTIAL).
FT CHAIN 21 931 NEUROFILIN-2.
FT DOMAIN 21 864 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 865 889 POTENTIAL.
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 28 142 CUB 1.
FT DOMAIN 149 267 CUB 2.

```

```
FT DOMAIN 277 427 F5/8 TYPE C 1.
FT DOMAIN 434 592 F5/8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DOMAIN 671 674 POLY-SER.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT CARBOHYD 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 809 813 MISSING (IN ISOCORM A17).
FT VARSPLIC 809 830 MISSING (IN ISOCORM A0).
FT CONFLICT 602 602 E -> K (IN REF. 1).
SQ SEQUENCE 931 AA; 104830 MM; 270CBAB69A0A797C CRC64;
```

Query Match 11.4%; Score 349; DB 1; Length 931;
Best Local Similarity 25.3%; Pred. No. 1.1e-18; Mismatches 166; Indels 84; Gaps 16;

```
DB 8 CGHLVYQDSQTMSTKNYPGTYPHNVCERTITPK-GKRLILRLG-DLDESQTCASDY 145
28 CGRLNKDAGYITSPGYPDYPSHONCEMIVYAPENOKIIVNFNFLEKHCKYDF 87
146 LLF-----TSSDOYGYPCGSMWPKELLNTSEYTFEESGSHISGRFLITY-----AS 196
88 IEIRDGDESEADLIGKHCNMT-APRTIISGSMLYIKFTDYAROGGFSRLYEIEFTGS 146
197 SD-----HFDLITC-----LERASHYLAKEYSF----- 220
147 EDCSKNTSPNGCTIESPGFPEKYPHNLDCFTTILAKRMETIIQLFLFDEHPLQYEG 206
221 -CPAGCDVADGIS--GNMVDGYRDTSLKCAIHAIGIADLGGQISVLQKRGISRYEG 277
207 DCKYMDIDIMDGIPHVGLIGKYCGTKPSRLSSTGILSLFTPTDMAVAK-----DG 259
278 ILANGVLSRQGSLSDKRFLELTSSNCCSRLSFE-----PDGQIRASSSSQSNESGDVHMS 333
260 FSAHYVYLVHDEPLENFQ-----CNVPLMESGRINDEQISASTY-----SDGRWT 305
334 PQGARLDQGPMSAGSDSNHMKPRELEIDLGKKKITGRTG--STOSNFFVYKSF 391
306 PQGSRRLGDDNGWTPNLDN-----KEYLDYDLRLTLMTALTAIYOGAISRETONGYVYKSY 361
392 VMNEKNNNSWKTYKGIIVNNEKVFQGSNFERDPVQNNFIPPIYARYVAVPQTHORIA 451
362 KLEVSTNGEDMMVYRHGKNH--KVFGQANNDATEVYLNKLHAPLLTFRVIRPQTHSGIA 419
452 LKVELLIGCOIT 462
420 LRLELFGCRVY 430
```

RESULT 3
NRP2_MOUSE STANDARD; PRT; 931 AA.
ID NRP2_MOUSE 035375; 035374; 035376; 035377; 035378;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
GN NRP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC STRAIN-BALB/C;
RX MEDLINE-97470888; PubMed-9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III";
RL Neuron 19:547-559(1997).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOTFORMS OF VEGF, AND THE PLGF-2 ISOTFORM OF PGF.
CC -1- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 6 ISOTFORMS: A22 (SHOWN HERE), A0, A5, A17,
CC B0 AND B5: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
CC NONNEURAL TISSUES INCLUDING LAMB BUDS, DEVELOPING BONES, MUSCLES,
CC INTERSTITIAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.
CC -1- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
CC IS DEVELOPMENTALLY REGULATED.
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AF022856; AAC53379.1; -
DR EMBL: AF022854; AAC53377.1; -
DR EMBL: AF022853; AAC53378.1; -
DR EMBL: AF022857; AAC53380.1; -
DR EMBL: AF022858; AAC53381.1; -
DR EMBL: AF022861; AAC53382.1; -
DR MGD: MGI:1100492; NRP2.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR000998; MAM.
DR Pfam: PF00754; CUB; 2.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR Pfam: PF00629; MAM; 1.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS50060; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 1 931
FT DOMAIN 21 864
FT TRANSMEM 865 889
FT DOMAIN 890 931
FT DOMAIN 28 142
FT DOMAIN 149 267
FT DOMAIN 277 427
FT DOMAIN 434 592
FT DOMAIN 642 802
FT DOMAIN 838 845
FT DISULFID 28 55
FT DISULFID 83 105
FT DISULFID 149 175
FT DISULFID 208 230
FT DISULFID 277 427
FT DISULFID 434 592
FT CARBOHYD 152 157
FT CARBOHYD 157 157

POTENTIAL.
NEUROFILIN-2.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
CUB 1.
CUB 2.
F5/8 TYPE C 1.
F5/8 TYPE C 2.
MAM.
POLY-SER.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).

QY 68 ALLAVSAPRLQAEELGDCGHLVYQDSGTMTSKNYPGTYPNHTYCEKTIIVPKG-KR 126
 11 ALPTTSLRALR-----SDKCGDTIKILSGVYLSPGYSTYHPQKCEMLQAPEPYQR 64
 QY 127 LILRLG-DLDESQTCASDYLLFTSSDQ-----YGPYCGSMTPYKELLMTSEVTFES 181
 65 IMINFNHPDLEDKCYDYVEVIDGDMAGRLMGKCYGKI-APPLVSSGPIPLFIKIVS 123
 QY 182 GSHISGGLFLTY-----ASS-----DHPDLITC----- 205
 DB 124 DETHGAGFISIRYEVERGECESNMTSSGMIKSPGFPKYPNSLCTYIIPAKMSEI 183
 QY 206 -LEKASHYLTKEYSKFCPCRPACRDVAGDISGNMYDGYRT-----SLCKAIHAGIINDEL 260
 DB 184 ILLEESFELPDSVTPGACRCRDLRI-----WDGPDVDPHIGRYC-GONNNGRVSST 238
 QY 261 GGOISVJQKRGISRYEGIILANGVLSRDGSLDKRFLFTSGCSKSLSFED---DQIRA 316
 DB 239 GILSMVYFDTSAIAKEGFSANYSQSVSEDFQ-----CMEPIGMESEIHSQITV 291
 QY 317 SSSMOSVNESGDDVHWSPPGARLDGCPASGSDSSNNHKPREMLEIDLGKKKTIQIRT 376
 DB 292 SSOYSAL-----WSESRKLYPBNGTPEBDS---VREMIQVLDGLLFFVSGIGT 339
 QY 377 TG--STQSNENFYVKSFFNMKNNSKMYKGIIVNNEKVFQSGNSFRDPYONNFPPI 434
 DB 340 OGAIKRETKKEYLYLKYTRVDVSSNGEDWILK--EGNKPVYFGNSNPIDVYIRPPKPV 397
 QY 435 VARYVRVYPOTMHORIALKVELIGCOIT 462
 DB 398 LTRFVRIRKPVSWENGVALREYGVCKIT 425

RESULT 5
 NRPI_HUMAN
 ID NRPI_HUMAN STANDARD; PRT; 923 AA.
 AC 014786; 060461.
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neupillin-1 precursor (Vascular endothelial cell growth factor 165 receptor).
 GN NRPI OR NRP OR VEGF165R.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
 RX MEDLINE=97433084; PubMed=9288753;
 RA He Z., Tessier-Lavigne M.;
 RT "Neupillin is a receptor for the axonal chemorepellent semaphorin III".
 RL Cell 90:739-751(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39. TISSUE-BREAST;
 RX MEDLINE=98188099; PubMed=9529250;
 RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
 RT "Neupillin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor".
 RL Cell 92:735-745(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (SOLUBLE/SNRPI ISOFORM), AND SEQUENCE OF 22-31. TISSUE-PROSTATIC ADENOCARCINOMA;
 RX MEDLINE=20183929; PubMed=10688880;
 RA Gagnon M.L., Bilenberg D.R., Gerechtman Z., Miao H.-Q., Takashima S., Soker S., Klagsbrun M.;
 RT "Identification of a natural soluble neupillin-1 that binds vascular endothelial growth factor: In vivo expression and antitumor activity".
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).

RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=20309748; PubMed=10748121;
 RA Gudman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
 RT "Neupillin-2 and neupillin-1 are receptors for the 165-amino acid form of vascular endothelial growth factor (VEGF) and of placenta growth factor-2, but only neupillin-2 functions as a receptor for the 145-amino acid form of VEGF".
 RL J. Biol. Chem. 275:18040-18045(2000).
 CC -1- FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS.
 CC -1- FUNCTION: THE SOLUBLE/SNRPI ISOFORM BINDS VEGF-165 AND APPEARS TO INHIBIT ITS BINDING TO CELLS. IT MAY ALSO INDUCE APOPTOSIS BY SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD VESSEL NUMBER AND INTEGRITY.
 CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN. THE SOLUBLE/SNRPI ISOFORM IS SECRETED.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: MEMBRANE-BOUND (SHOWN HERE) AND SOLUBLE/SNRPI. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND SOLUBLE/SNRPI ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT TISSUES. IN THE DEVELOPING EMBRYO IT IS FOUND PREDOMINANTLY IN THE NERVOUS SYSTEM. IN ADULT TISSUES, IT IS HIGHLY EXPRESSED IN HEART AND PLACENTA. MODERATELY IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS. AND LOW IN ADULT BRAIN. THE SOLUBLE/SNRPI ISOFORM IS FOUND IN LIVER HEPATOCYTES, KIDNEY DISTAL AND PROXIMAL TUBULES.
 CC -1- SIMILARITY: BELONGS TO THE NEURILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AF018956; AAC51759.1; -
 CC DR EMBL, AF016050; AAC12921.1; -
 CC DR EMBL, AF145712; AAP44344.1; -
 CC MIM, 602069;
 CC InterPro: IPR000859; CUB.
 CC InterPro: IPR000421; FA58_C.
 CC InterPro: IPR001092; HLH_dlm.
 CC InterPro: IPR000998; MAM.
 CC Pfam, PF00431; CUB; 2.
 CC Pfam, PF00754; F5_F8_type_C; 2.
 CC Pfam, PF00629; MAM; 1.
 CC PRINTS, PR00020; MAMDOMAIN.
 CC SMART, SM00042; CUB; 2.
 CC SMART, SM00231; FA58C; 2.
 CC SMART, SM00137; MAM; 1.
 CC PROSITE, PS01180; CUB; 2.
 CC PROSITE, PS01285; FA58C_1; 2.
 CC PROSITE, PS01286; FA58C_2; 2.
 CC PROSITE, PS00740; MAM_1; 1.
 CC PROSITE, PS00600; MAM_2; 1.
 CC Transmembrane: Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Alternative splicing.
 FT SIGNAL 1 21
 FT CHAIN 22 923 NEURILIN-1.
 FT DOMAIN 22 856 EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 857 879 POTENTIAL.
FT DOMAIN 880 923 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 141 CUB 1.
FT DOMAIN 147 265 CUB 2.
FT DOMAIN 275 424 F5/8 TYPE C 1.
FT DOMAIN 431 583 F5/8 TYPE C 2.
FT DOMAIN 645 811 MAM.
FT DISULFID 27 54 PROBABLE.
FT DISULFID 82 104 PROBABLE.
FT DISULFID 147 173 PROBABLE.
FT DISULFID 206 228 PROBABLE.
FT DISULFID 275 424 BY SIMILARITY.
FT DISULFID 431 583 BY SIMILARITY.
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 842 842 EEP -> GIK (IN SOLUBLE/SNRPI ISOFORM).
FT VARSPIC 642 644 MISSING (IN SOLUBLE/SNRPI ISOFORM).
FT CONFLICT 26 26 K -> E (IN REF. 1).
FT CONFLICT 749 749 D -> H (IN REF. 2).
FT CONFLICT 855 855 E -> D (IN REF. 2).
SQ SEQUENCE 923 AA: 103120 MW: ADEADCA4849E5D57 CRC64:

```

Query Match 11.0%; Score 335; DB 1; Length 923;
 Best Local Similarity 24.9%; Pred. No. 1,3e-17;
 Matches 124; Conservative 76; Mismatches 176; Indels 122; Gaps 23;

```

QY 64 REL-----LALLAVSAPRLQAEELGDCGHLVYQDSGTMTSKNPPYNNHTVCEK 117
DB 3 RELPLCAVLAVALVAPAGAFR-----NDKCDTIKIESPYLTSPGYSHYHPEKCEW 56
QY 118 TTVPEKG-KRLILRG-DIDISQTCASDYL-LFTSSSDQ---YPPYCGSMVPELLIN 171
DB 57 LQAADPPYRIMINPNPHDLEDROCKYDYVEFDGEMNGHFGRKFGCKT-APPYVSS. 115
QY 172 TSEVTVREFSGSHIGRGLTYA-----SSDHPDLITC-- 205
DB 116 GPELFKEVSDYETHGACPSIRYEIFKRQPECSQNTYTPSGYIKSPGPEKYPNLECY 175
QY 206 -----LEASHYIKTEYSKFCPRG--CRVAGDISNMVDGYDTSLCLKAAH. 252
DB 176 IYFAPKASEIILEFSEFDLEPDSNP--PEGMFCRDREI-----WDGEPD-----VGP 223
QY 253 AG-IIADELGOI-----SYLORKGISRYEGILANGVLSRDGSLDKRFLF. 297
DB 224 IGRYGGKTPRRIRSSSGILSMVFTDLSIAKEGFSANYSVQSSV-SDEPK----- 274
QY 298 TENGCSRSLSFEP---DQIRASSSMOSVNESDQVHWSPGQARLQDQGPSWASGDSSN 353
DB 275 ---CMEALGMESGEIHSQITASSQYST-----NMSABERSRLNYPENCMTGEDESY 322
QY 354 NKKPREMLEIDEGEKKTKIGRTTG--SNQSNFNFYVKSFEVNFNNNSKKKTYGIYVN 411
DB 323 ---REMIQVDDGLKLFYAVATQGISKEYKKKTYKTKIDVSSNGEDWITIK--EGN 376
QY 412 EEKVFQGSNFRDPYONNFIPIPIYAVYVVPOTWGORIALKVELLIGCOIT-----QGNDS 467
DB 377 KVLPLQGNMNPIDVVAVPEPKLITRFVIRKIPATMETGISMFEYVGCKITIDYPCSGMIG 436
QY 468 LVMKTSOSTSVSTYKED 485
DB 437 MVSGLISDSQITSSNOGD 454

```

```

RESULT 6
ID NRPI_RAT STANDARD: PRT: 922 AA.
AC Q9QWJ9;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuregulin-1 precursor (Vascular endothelial cell growth factor 165
DE receptor).
GN NRPI.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRPAGUE-DWILEY;
RX MEDLINE=97433085; PubMed=9288754;
RA Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,
RA Ginty D.D.;
RT "Neuregulin is a semaphorin III receptor.";
RL Cell 90:753-762(1997).
CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC CERTAIN NEURONAL CIRCUITS AND IN ORGANORENESIS OUTSIDE THE NERVOUS
CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
CC BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165
CC ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
CC INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
CC IF MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- TISSUE SPECIFICITY: FOUND IN THE EMBRYONIC NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF016296; AAC53337.1; -.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000421; FA58.C.
DR InterPro; IPR001092; HTH_dtm.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR TRANSFAM; PS50060; MAM_2; 1.
RW Transmembrane; Glycoprotein;
FT SIGNAL 1 21
FT CHAIN 22 922
FT DOMAIN 22 855
FT TRANSMEM 856 880
FT TRANSMEM 881 922
FT DOMAIN 27 141
FT DOMAIN 147 265
FT DOMAIN 275 424
FT DOMAIN 431 583
FT DOMAIN 645 811
FT DISULFID 27 54
FT DISULFID 82 104
FT DISULFID 147 173
FT DISULFID 206 228
FT DISULFID 275 424
FT DISULFID 431 583

```

RESULT	7
FA5_BOVIN	
ID	FA5_BOVIN
AC	Q28107; STANDARD;
D7	Q28107; Q28108; PRT; 221 AA.
D7	01-NOV-1997 (Rel. 35, Created)
D7	01-NOV-1997 (Rel. 35, Last sequence update)
D7	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Coagulation factor V precursor (Activated protein C cofactor).
GN	F5.
OS	Bos taurus (Bovine).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC	Bovidae; Bovinae; Bos.
OX	NCBI_TaxID=9913;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	TISSUE=liver;
RX	MEDLINE=92147638; PubMed=1737753;
RA	Giunto E.R., Esmon C.T., Mann K.G., Macgillivray R.T.;
RT	"The complete cDNA sequence of bovine coagulation factor V.";
TL	J. Biol. Chem. 267:2971-2978(1992).
CC	- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC	- SUBUNIT: Factor Va is composed of a heavy chain and a light chain, noncovalently bound. The interaction between the two chains is calcium-dependent.
CC	- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14 AA repeats.

CC	-1-	P1M Thrombin activates factor V proteolytically to the active cofactor, factor Va (formation of a heavy chain at the N-terminus and a light chain at the C-terminus).
CC	-1-	P1M: SUFFICIENT IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).
CC	-1-	SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF 2 PLASTOCYANIN-LIKE REPEATS.
CC	-1-	SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-1-	SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
CC		-----
DR	EMBL:	M81440; AAA30512.1; -
DR	EMBL:	M81441; AAA30513.1; -
DR	HSSP:	P00450; 1KCW.
DR	InterPro:	IPR001117; Cu-oxidase.
DR	InterPro:	IPR000421; FA5B_C.
DR	Ffam:	PF00394; Cu-oxidase; 3.
DR	Ffam:	PF00754; F5_F8_type_C; 2.
DR	SMART:	SM00231; FA5Bc; 2.
DR	PROSITE:	PS00079; MULTICOPPER_OXIDASE1; 2.
DR	PROSITE:	PS01285; FA5Bc_1; 2.
KW	PROSITE:	PS01286; FA5Bc_2; 2.
KW	Blood coagulation;	Glycoprotein; Sulfation; Calcium; Signal; Zymogen; Repeat.
KW	SIGNAL.	
FT	CHAIN	1 28 POTENTIAL.
FT	CHAIN	29 2211 COAGULATION FACTOR V.
FT	CHAIN	29 741 HEAVY CHAIN (BY SIMILARITY).
FT	PEPTIDE	742 1564 ACTIVATION PEPTIDE (CONNECTING REGION) (BY SIMILARITY).
FT	DOMAIN	1565 2211 LIGHT CHAIN (BY SIMILARITY).
FT	DOMAIN	30 327 F5/8 TYPE A 1.
FT	DOMAIN	30 193 PLASTOCYANIN-LIKE 1.
FT	DOMAIN	203 327 PLASTOCYANIN-LIKE 2.
FT	DOMAIN	348 686 F5/8 TYPE A 2.
FT	DOMAIN	348 525 PLASTOCYANIN-LIKE 3.
FT	DOMAIN	535 686 PLASTOCYANIN-LIKE 4.
FT	DOMAIN	696 1564 B.
FT	SIMILAR	899 915 TO 17 AA REPEATS IN HUMAN FAS.
FT	DOMAIN	1124 1151 2 X 14 AA TANDEM REPEATS.
FT	REPEAT	1124 1137 1-1.
FT	REPEAT	1138 1151 1-2.
FT	DOMAIN	1188 1453 30 X 9 AA APPROXIMATE TANDEM REPEATS OF [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
FT	REPEAT	1188 1196 2-1.
FT	REPEAT	1197 1205 2-2.
FT	REPEAT	1206 1214 2-3.
FT	REPEAT	1215 1223 2-4.
FT	REPEAT	1224 1232 2-5.
FT	REPEAT	1233 1241 2-6.
FT	REPEAT	1242 1250 2-7.
FT	REPEAT	1251 1259 2-8.
FT	REPEAT	1260 1268 2-9.
FT	REPEAT	1269 1277 2-10.
FT	REPEAT	1278 1286 2-11.
FT	REPEAT	1287 1295 2-12.
FT	REPEAT	1296 1304 2-13.
FT	REPEAT	1305 1313 2-14.
FT	REPEAT	1314 1322 2-15.
FT	REPEAT	1323 1331 2-16.
FT	REPEAT	1332 1340 2-17.
FT	REPEAT	1341 1349 2-18.
FT	REPEAT	1350 1358 2-19.
FT	REPEAT	1359 1367 2-20.
FT	REPEAT	1368 1376 2-21.
FT	REPEAT	1377 1385 2-22.
FT	REPEAT	1386 1394 2-23.

```

FT REPEAT 1395 1403 2-24.
FT REPEAT 1404 1412 2-25.
FT REPEAT 1413 1421 2-26.
FT REPEAT 1422 1430 2-27.
FT REPEAT 1431 1439 2-28.
FT REPEAT 1440 1444 2-29. (PARTIAL).
FT REPEAT 1445 1453 2-30.
FT DOMAIN 1569 1890 F5/8 TYPE A 3.
FT DOMAIN 1569 1738 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1894 2048 F5/8 TYPE C 1.
FT DOMAIN 2053 2208 F5/8 TYPE C 2.
FT SITE 741 742 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1564 1565 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT DISULFID 167 193 PROBABLE.
FT DISULFID 499 525 PROBABLE.
FT DISULFID 1712 1738 PROBABLE.
FT DISULFID 1894 2048 BY SIMILARITY.
FT DISULFID 2053 2208 BY SIMILARITY.
FT MOD_RES 697 697 SULFATION (POTENTIAL).
FT MOD_RES 701 701 SULFATION (POTENTIAL).
FT MOD_RES 730 730 SULFATION (POTENTIAL).
FT MOD_RES 1513 1513 SULFATION (POTENTIAL).
FT MOD_RES 1529 1529 SULFATION (POTENTIAL).
FT MOD_RES 1537 1537 SULFATION (POTENTIAL).
FT MOD_RES 1541 1541 SULFATION (POTENTIAL).
FT MOD_RES 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 587 592 N-TERMINAL (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;

```

Query Match 10.8%; Score 331; DB 1; Length 2211;
 Best Local Similarity 32.5%; Pred. No. 9.2e-17;
 Matches 88; Conservative 48; Mismatches 89; Indels 46; Gaps 10;

```

QY 209 ASHLYKTEY-SKCPA-----GCRDVAGDISGNM--DGYRDTSLCKRAAIHAGIIDE 259
DB 1968 AKHLIKYTERFCVAISLDRKKMRFKGSSTNVMYFGSNDASITIKENOIDPVA-- 2025
QY 260 LGGQISVLRKGISRYGILANGVLSGSLSKRLFTSNGCSKSLSE----PDGQIR 315
DB 2026 -----RYIRISPTGSYNKPALRLLEQGC-EVNGCSTPIAGSESGKIEKKQIT 2070
QY 316 ASS-----SMQSVNESGDVHMSPGQARLDQGG--PSNASGSSNNHNPRLMLEIDGAEKK 370
DB 2071 ASSFKKSMWG-----NYMEPLRLINAGGRVANAOKANNNN-----OMLQIDLIKIKK 2119

```

```

QY 371 ITGIRTTGTSQSNENFYKSPVAMFKNNNSKMTKYGIYNNNEKVPQGSNFDPOYNNF 430
DB 2120 ITAIVTQCKSLSESMYKSYTHYSQGDWPKYREKSSMWKIFEGANNVGHVKNFF 2179
QY 431 IPIVARYRVVPTQTHORIALKVELIGCOI 461
DB 2180 NPPIISIRIRIIPKTMNSIALRLFLGCGM 2210

RESULT 8
ID NPPI_XENLA STANDARD; PRT; 928 AA.
AC P28824;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neutrophilin-1 precursor (A5 protein) (A5 antigen).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OX Xenopodinae; Xenopus.
NCBI_TaxID=8353;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=91337458; PubMed=1908252;
RA Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;
RT "The A5 antigen, a candidate for the neuronal recognition molecule,
RL Neuron 7:295-307(1991).
CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
CC SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
CC BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
CC NEURONS.
CC -1- SIMILARITY: BELONGS TO THE NEUTROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.jab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: D10467; BAA01260.1; -
DR InterPro: IPR000859; CUB; 1.
DR InterPro: IPR000421; FAS8_C.
DR InterPro: IPR000958; MAM.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PRO0020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FAS8C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FAS8C_1; 2.
DR PROSITE: PS01286; FAS8C_2; 2.
DR PROSITE: PS00600; MAM_2; 1.
DR Transmembrane; Glycoprotein; Neuropeptide; Signal; Receptor;
KW Antigen.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 928 NEUTROFILIN-1.

```



```

FT DOMAIN 22 860 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 861 883 POTENTIAL.
FT DOMAIN 884 928 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 141 CUB 1.
FT DOMAIN 147 265 CUB 2.
FT DOMAIN 275 424 F5/8 TYPE C 1.
FT DOMAIN 431 584 F5/8 TYPE C 2.
FT DOMAIN 646 812 MAM.
FT DISULFID 27 54 PROBABLE.
FT DISULFID 82 104 PROBABLE.
FT DISULFID 147 173 PROBABLE.
FT DISULFID 206 228 PROBABLE.
FT DISULFID 275 424 BY SIMILARITY.
FT DISULFID 431 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 928 AA; 103416 MW; AF6B323B0AC789D CRC64;

```

Query Match Best Local Similarity 10.8%; Score 330; DB 1; Length 928;

Matches 112; Conservative 74; Mismatches 171; Indels 106; Gaps 20;

```

QY 66 LIALLLAVSAPRLQAEELGDCGHLVYQDSGMTSKNPTGCTVNTTVEKTTVKG- 124
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 5 LLSGCCMLCLSRSSMSNRKDCDTIKTSPSYLSAGPHSPQRCWMLQAEVHY 64

QY 125 KRLLRLUG-DIDIESQTCADSYLLF---TSSDQYPCYCSMTVPKELLINTSEVTVRF 179
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 65 QRININPNPHEDLEDRCKDYVEVIDGNANGLCKYCGKI-APSPVSTGSPITRF 123

QY 180 ESGHISGRGLFLY-----ASSD-----HPDLITC----- 205
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 124 VSDYETGAGTSIRYEVFKGPECCSRNFTSNVGIKPKYPEKYPNALCTEYIIIFAPKMQ 183

QY 206 ---LERASHYLTKEYSKFCFAG--CR-DVAGDISGNVNDYRDTSLCKAAIHAGIAD 259
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 184 EIVLEFESFELEADSN--AVGGGTCRDWLG-----IMDFPFG-----VGRH 223

QY 260 LGGGIVSLQKKGISRYGIIA-----NGVLSRDG-----SLDKRFLFTSNGCSRS 305
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 224 IGRYCGONTGVRVNSFGILSMIFHTDSALAKGFFANFSVVSNTDEDE---CCKEA 278

QY 306 LSEFPD---GQIRASSSQSVNESGDVHMSFGOARLDQGPMSAGDSNNHKKPREWL 361
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 279 LGMESGRIHFDQISVSQY-----SMWMSAERSKLNLYVENGWTPGEDT---YKEWI 326

QY 362 EIDLGKKKITTGRTTG--STQSNFNFYVKSFYVNFKNNSKWKTYKGIYVNEEKYFOGN 419
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 327 QVDLENIRFVSGIGTQGAISKEFKKYFYKVDISSNEDWITLKD--GNHILVFTGN 384

QY 420 SNRPDPYONNFPIPIVARYRVVVPOTMHOQIAKVELIGQIT 462
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 385 TDAITDVYVRFPSKPVITRFVRLKPTWENGISLRFELYGCKIT 427

RESULT 9
ID NRPI_MOUSE STANDARD; PRT; 923 AA.
AC P97333;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-1 precursor (A5 protein).
GN NRPI OR NRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=Embryonic brain;
RX MEDLINE=96353149; Pubmed=8748368;
RA Kawakami A., Katsukawa T., Takagi S., Fujisawa H.;
RT "Developmentally regulated expression of a cell surface protein,
RL J. Neurobiol. 29:1-17(1996).
CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
CC BINDS TO SEMAPHORIN 3A, THE PLEK-2 ISOFORM OF PEF, THE VEGF-165
CC INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
CC IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE NEUROTROPHIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; D50086; BAA08769.1; -
DR MGD; MGI:106206; NRP.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000421; FASB_C.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FASB_C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FASB_C; 1; 2.
DR PROSITE; PS01286; FASB_C; 2.
DR PROSITE; PS00740; MAM; 1; 1.
DR PROSITE; PS50060; MAM; 2; 1.
KW Transmembrane; Glycoprotein;
FT SIGNAL 1 21
FT CHAIN 22 923
FT DOMAIN 22 856
FT TRANSMEM 857 879
FT DOMAIN 880 923
FT DOMAIN 27 141
FT DOMAIN 147 265
FT DOMAIN 275 424
FT DOMAIN 431 583
FT DOMAIN 645 811
FT DISULFID 27 54
FT DISULFID 82 104
FT DISULFID 147 173
FT DISULFID 206 228
FT DISULFID 275 424
FT DISULFID 431 583
FT CARBOHYD 150 150
FT CARBOHYD 261 261
FT CARBOHYD 300 300
FT CARBOHYD 522 522
FT CARBOHYD 842 842
SQ SEQUENCE 923 AA; 103020 MW; 0644B8A170796808 CRC64;

```

Query Match Best Local Similarity 10.8%; Score 329; DB 1; Length 923;

Matches 112; Conservative 74; Mismatches 171; Indels 106; Gaps 20;

Matches 113; Conservative 79; Mismatches 161; Indels 118; Gaps 21;

OY 64 RGL-----TALLAVSAPLRLQAEELGCGGHLVTTODSGTMTSKNPGYPMHTYCEK 117
 DB 3 RGLPLCATLALALALALAFR-----SDKCGGTIKIENPGLYSPGPHSYHSEKCEW 56
 OY 118 TTPVPRG-KRLILRLRG-DLDESQTCASDYLLFTSSDQ-----YGPYCGSMTVREKELLIN 171
 DB 57 LIQAPEPYQRIIINPNFDEEDRDCKDYEVLDGENBGRMLGKCGCKI-APSPVSS 115
 OY 172 TSEVTVREGSGHISGRGELLTYA-----SSDHPDLITC-- 205
 DB 116 GPFLFKFVSDYETHGAFSIRYEIFKRGPCSONYTPATGVIKSPGPERYPMCLDCTY 175
 OY 206 -----LERASHLTKTEYSFCFAG--CDVADIDIGNVNDGQRODSLLOKKAH 252
 DB 176 IIFAPKMEIILFEFSDLEDSDNP--PGGMFCRIDLET---WDGPE-----VGP 223
 OY 253 AG-IADLAGQI-----SVLORKGISREGLIANGVLSRDSLSKRLFL 297
 DB 224 IGRVCGKTPGRIRSSGVLNMFYTDALAKEGFSANYSVLQSSI-SEDK----- 274
 OY 298 TSNCGSRSLSEFP---DQIRASSWQSVNESGDQVHWSFGQARLDQGSMASSGSSN 353
 DB 275 ---CMEALGMEGSEIHSDDITASSQYGT-----NMSYERSKLNPNKMGWTPGEDSY 322
 OY 354 NHPREWEIDLGKKKKTGTGIRTTG--STQSNFNFYKSFPMFNKNNNSKMTKGIYNN 411
 DB 323 ---KEMLOVDLGLRLFTAVGTQALSKERKKKYVTVYVDISSNEDMISLK--EGN 376
 OY 412 EEKVEGNSNFRDPYNNFPIPIYARVYVVPQTHORIALKVELIGQIT 462
 DB 377 KALIFOGNTPDVLGVFSKPLTRFVRIRKPSWETGISMRFEVYCKIT 427

RESULT 10
 FAS_HUMAN STANDARD; PRT; 2224 AA.
 ID FAS_HUMAN
 AC P12259; Q14285;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE FROM N.A.
 RX MEDLINE-92232668; PubMed-1567832;
 RA Cripe L.D., Moore K.D., Kane W.H.;
 RT "Structure of the gene for human coagulation factor V.";
 RL Biochemistry 31:377-3785(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87260886; PubMed-3110773;
 RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 RN [3]
 RP SEQUENCE OF 1-1600 FROM N.A.
 RX MEDLINE-88107560; PubMed-2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 RT region of human factor V, a blood coagulation factor with four types
 RT of internal repeats.";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE-86313665; PubMed-3092220;
 RA Kane W.H., Davie E.W.;

RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RC TISSUE-FIBROBLAST;
 RX MEDLINE-93203619; PubMed-8454869;
 RA Shen N.L.L., Fan S.-T., Pyatt J., Graf R., Lapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes.";
 RL J. Immunol. 150:2992-3001(1993).
 RN [6]
 RP SULFATION.
 RX MEDLINE-94264012; PubMed-8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Seligshohn U.,
 RA Kaufman R.J.,
 RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity.";
 RL Biochemistry 33:6952-6959(1994).
 RN [7]
 RP SULFATION.
 RX MEDLINE-80366699; PubMed-2168225;
 RA Horlin G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V.";
 RL Blood 76:946-952(1990).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE-20052169; PubMed-10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Orel T.L., Bourenkov G.P., Bartunk H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V.";
 RL Nature 402:434-439(1999).
 RN [9]
 RP VARIANT APCR GLN-534.
 RX MEDLINE-94217810; PubMed-8164741;
 RA Bertina R.M., Koelme B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reijlma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C.";
 RL Nature 369:64-67(1994).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 CC REPEATS.
 CC -1- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
 CC -1- DISEASE: OREN PARAHEMOPHILIA, AN HEMORRHAGIC DISASTERS, IS DUE
 CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
 CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
 CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
 CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
 CC IMPLANTATION.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG. TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; L32779; AAB59401.1; JOINED.
DR EMBL; L32755; AAB59401.1; JOINED.
DR EMBL; L32756; AAB59401.1; JOINED.
DR EMBL; L32757; AAB59401.1; JOINED.
DR EMBL; L32758; AAB59401.1; JOINED.
DR EMBL; L32759; AAB59401.1; JOINED.
DR EMBL; L32760; AAB59401.1; JOINED.
DR EMBL; L32761; AAB59401.1; JOINED.
DR EMBL; L32762; AAB59401.1; JOINED.
DR EMBL; L32763; AAB59401.1; JOINED.
DR EMBL; L32764; AAB59401.1; JOINED.
DR EMBL; L32765; AAB59401.1; JOINED.
DR EMBL; L32766; AAB59401.1; JOINED.
DR EMBL; L32767; AAB59401.1; JOINED.
DR EMBL; L32768; AAB59401.1; JOINED.
DR EMBL; L32769; AAB59401.1; JOINED.
DR EMBL; L32770; AAB59401.1; JOINED.
DR EMBL; L32771; AAB59401.1; JOINED.
DR EMBL; L32772; AAB59401.1; JOINED.
DR EMBL; L32773; AAB59401.1; JOINED.
DR EMBL; L32774; AAB59401.1; JOINED.
DR EMBL; L32775; AAB59401.1; JOINED.
DR EMBL; L32776; AAB59401.1; JOINED.
DR EMBL; L32777; AAB59401.1; JOINED.
DR EMBL; L32778; AAB59401.1; JOINED.
DR EMBL; M16967; AAB59401.1; JOINED.
DR EMBL; M14335; AAB59532.1; -.
DR PIR; A25897; A25897.
DR PIR; A28028; A28028.
DR PDB; 1CZS; 26-NOV-99.
DR PDB; 1CZT; 26-NOV-99.
DR PDB; 1CZV; 26-NOV-99.
DR MIM; 134400; -.
DR MIM; 188055; -.
DR MIM; 227310; -.
DR MIM; 227400; -.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA5B_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR SMART; SM00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA5B_C; 2.
DR PROSITE; PS00079; MULTICOPPEL_OXIDASE1; 2.
DR PROSITE; PS01285; FA5B_C; 2.
DR PROSITE; PS01286; FA5B_C; 2.
DR Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat; Polymorphism; Disease mutation; Thrombophilia; 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 2224
FT CHAIN 29 737
FT PEPTIDE 738 1573
FT CHAIN 1574 2224
FT DOMAIN 30 329
FT DOMAIN 30 193
FT DOMAIN 203 329
FT DOMAIN 348 684
FT DOMAIN 348 526
FT DOMAIN 536 684
FT DOMAIN 692 1573
FT DOMAIN 895 928
FT REPEAT 895 911
FT REPEAT 912 928
FT SIMILAR 1135 1148
FT DOMAIN 1185 1501
FT REPEAT 1185 1193
FT REPEAT 1194 1202
FT REPEAT 1203 1211
FT REPEAT 1212 1220
FT REPEAT 1221 1229
FT REPEAT 1230 1238
FT REPEAT 1239 1247
FT REPEAT 1248 1256

```

```

FT REPEAT 1257 1265 2-9.
FT REPEAT 1266 1274 2-10.
FT REPEAT 1275 1283 2-11.
FT REPEAT 1284 1292 2-12.
FT REPEAT 1293 1301 2-13.
FT REPEAT 1302 1310 2-14.
FT REPEAT 1311 1319 2-15.
FT REPEAT 1320 1328 2-16.
FT REPEAT 1329 1337 2-17.
FT REPEAT 1338 1346 2-18.
FT REPEAT 1347 1355 2-19.
FT REPEAT 1356 1364 2-20.
FT REPEAT 1365 1373 2-21.
FT REPEAT 1374 1382 2-22.
FT REPEAT 1383 1391 2-23.
FT REPEAT 1392 1400 2-24.
FT REPEAT 1401 1409 2-25.
FT REPEAT 1410 1418 2-26.

```

Query Match

10.4%; Score 317; DB 1; Length 2224;

Best Local Similarity 42.0%; Pred. No. 1, le-15;

Matches 71; Conservative 30; Mismatches 52; Indels 16; Gaps 5;

```

QY 300 NCGSRSLSEF-----PDGQIRASS---SMQSVNDSGPOVHMSPCQARTLQDGPMSAGDSS 352
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2064 NCGSTPLGMNGKIEKNIENKQITASSFKSW-----GD--YWEPRATLNAQGRVMAQAKA 2116
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 353 NNHKPREWLEIDGEEKKIKITGTTGSGTOSNENFYKSFVMEKNKNSKTKYGVNNE 412
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2117 NNKK--CWLEIDILKIKITVAITTOCKSLSSMEYKSYTIHSEGVGEMKPYRLKSSWV 2174
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 413 EKVEFGNSNFRDPVONNFTPIVARYRVYVPOVHRIKLYLJCCQI 461
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2175 DKIFEGNTNKGHVKNFNPITSRFIRVLPKTNQSTITRLRLFGCDI 2223
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 11

```

FA5_PIG STANDARD; PRT; 2258 AA.
ID FA5_PIG
AC O9GRPL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C
DE cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS.
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
RT protein modeling of membrane binding sites and comparative anatomy of
RT domains.";
RL Cell. Mol. Life Sci. 58:148-159(2001).
CC -I- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -I- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -I- DOMAIN: Domain B contains 41 x 9 AA tandem repeats. Domains C1
CC and C2 may be involved in membrane binding.
CC -I- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.

```


RX MEDLINE-91093266; PubMed-1898735;
 RA Leye A., van Schijndel H.B., Niehrs C., Hutner W.B., Verbeet M.P.,
 RA Meyers K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor.";
 RL J. Biol. Chem. 266:740-746(1991).
 RN [7]
 RP SULFATION.
 RX MEDLINE-92207952; PubMed-1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII.";
 RL Biochemistry 31:3315-3325(1992).
 RN [8]
 RP STRUCTURE BY NMR OF 2322-2343
 RX MEDLINE-95200924; PubMed-7893714;
 RA Gilbert G.E., Baleja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL Biochemistry 34:3022-3031(1995).
 RN [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-91221499; PubMed-1902642;
 RA Gitschler J.;
 RT "The molecular basis of hemophilia A.";
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-89088506; PubMed-2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL Blood 73:1-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-95245332; PubMed-7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RP VARIANT GLN-2326.
 RX MEDLINE-86235434; PubMed-3012775;
 RA Gitschler J., Wood W.I., Shuman W.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophilic.";
 RL Science 232:1415-1416(1986).
 RN [13]
 RP VARIANT PRO-2135.
 RX MEDLINE-88096539; PubMed-3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschler J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RP VARIANT GLN-2228.
 RX MEDLINE-88191889; PubMed-2833855;
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RP VARIANT GLY-291.
 RX MEDLINE-88220354; PubMed-2835904;
 RA Youssoufian H., Wong C., Aronis S., Platakoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT CYS-1708.
 RX MEDLINE-89274393; PubMed-2499363;
 RA O'Brien D.P., Tuddenham E.G.;

RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT CYS-391.
 RX MEDLINE-90001543; PubMed-2506948;
 RA Shima M., Ware J., Yoshitaka A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT LEU-189.
 RX MEDLINE-90057680; PubMed-2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A.";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT LEU-2326.
 RX MEDLINE-89197216; PubMed-2495245;
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene.";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HIS-391.
 RX MEDLINE-89264602; PubMed-2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT CYS-1708.
 RX MEDLINE-90105723; PubMed-2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANTS GLN-2228 AND LEU-2326.
 RX MEDLINE-90123183; PubMed-2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT CYS-391.
 RX MEDLINE-90329422; PubMed-1973901;
 RA Patlison J.K., McVey J.H., Boon M., Afari A., Tuddenham E.G.;
 RT "CRM+ haemophilia A due to a missense mutation (372->Cys) at the
 RT internal heavy chain thrombin cleavage site.";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS PHE-1699 AND CYS-1708.
 RX MEDLINE-90152691; PubMed-2105906;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE-90169988; PubMed-2106480;
 RA Treysman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene.";

Tue Jul 9 08:27:37 2002

us-09-691-344a-4.rsp

Page 17

Dy 402 MKTKGIVNNEEKYFQGSNFRDPVQNNEIPRIYARYVWPQTWNRIALKVELIGC 459
+ : : : : :
Db 408 WTIVE--EGSSSKVFQGLDNNSHKNINIEKKPFMARUYRVLPVSNNHNTIURLLEIGC 463

Search completed: July 9, 2002, 02:26:11
Job time: 648 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 9, 2002, 02:14:03 ; Search time 130.73 Seconds

(without alignments)
775,453 Million cell updates/sec

Title: US-09-691-344A-4
Perfect score: 3052
Sequence: 1 MGFGAGQRLRPVAPRSSAE.....YDNKEMQKLDLITSDMAG 586

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1310.5	42.9	503	11	Q9D4J3 mus musculus
2	1154.5	37.8	460	11	Q9D696 mus musculus
3	942	30.9	769	11	Q91ZV2 rat
4	937	30.7	769	11	Q91ZV3 mus musculus
5	929	30.4	775	4	Q96PD2 mus musculus
6	572.5	18.8	364	4	Q14089 homo sapien
7	349	11.4	555	4	Q9H2E2 homo sapien
8	349	11.4	901	4	Q9H2E4 homo sapien
9	349	11.4	901	4	Q9H2D5 homo sapien
10	349	11.4	906	4	Q9H2D4 homo sapien
11	349	11.0	609	4	Q96190 homo sapien
12	335	11.0	609	4	Q96190 homo sapien
13	335	11.0	609	4	Q96190 homo sapien
14	335	11.0	609	4	Q96190 homo sapien
15	325.5	10.7	921	11	Q90X38 rat
16	321	10.5	2224	4	Q43737 homo sapien

17	318	10.4	2193	11	Q88783 mus musculus
18	306.5	10.0	2343	6	O18806 canis famill
19	305.5	10.0	2343	6	O62730 canis famill
20	299.5	9.8	779	4	Q9NTT3
21	290	9.5	426	11	Q9WTS3
22	285.5	9.4	216	4	Q14286
23	279.5	9.2	453	11	Q9R1X9
24	277	9.1	2119	13	Q90X47
25	268	8.8	480	4	O43854
26	264	8.7	480	11	O35474
27	238.5	7.8	363	6	O77718
28	226	7.4	1128	11	Q88442
29	221.5	7.3	224	11	Q9R1M6
30	217	7.1	1158	4	O14113
31	216.5	7.1	719	11	Q61281
32	209.5	6.9	728	6	O97567
33	199.5	6.5	858	5	O76470
34	197	6.5	721	4	O96M80
35	194.5	6.4	335	4	O9RT19
36	188.5	6.2	3843	5	Q9U5D0
37	187.5	6.1	3843	5	O9U094
38	186.5	6.1	745	4	O96M75
39	185.5	6.1	963	4	O96M57
40	182.5	6.0	764	11	Q9D2L5
41	182.5	6.0	764	11	O54860
42	182	6.0	604	4	Q96M00
43	174	5.7	1524	13	Q91674
44	173.5	5.7	926	4	Q90Q00
45	173	5.7	414	11	Q9CX06

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	503 AA.
ID	Q9D4J3			
AC	Q9D4J3			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DIC-2001 (TREMBLrel. 19, Last annotation update)			
DE	4631413K1RIK PROTEIN.			
GN	4631413K1RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda I.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Koehle H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,			
RA	Blake J., Offelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guernicich S., Hill D., Hofmann M., Hume D.A., Kamlay M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombert P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	*Functional annotation of a full-length mouse cDNA collection.*;			
RL	Nature 409:685-690(2001).			
CC	- SIMILARITY: CONTAINS 1 CUB DOMAIN.			
DR	EMBL; AK016485; BAB30265.1; -			

DR MGD: MGI:1913936; 4631413K11RLK.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR004043; LCC1.
 DR Pfam: PF00431; CUB: 1.
 DR SMART: SM00042; CUB: 1.
 DR PROSITE: PS01180; CUB: 1.
 DR SEQUENCE 503 AA; 54547 MW; FEB121E845CAA06B CRC64;

Query Match
 Best Local Similarity 42.9%; Score 1310.5; DB 11; Length 503;
 Matches 268; Conservative 26; Mismatches 33; Indels 199; Gaps 1;

QY 61 AAGRGILLALLVAPRLQAELEGCGHLYTODSGTMTSKNPGTYPNHTECKTT 120
 DB 5 AAGPSVALLFAVCAPLRLQAELEGCGHLYTODSGTMTSKNPGTYPNHTECKIT 64
 QY 121 VKKGRLLIRLDDLESTQCSADYLLFTSSDQGYPCGSMVTKRELLNTSEVTRPE 180
 DB 65 VKKGRLLIRLDDLESTQCSADYLLFTSSDQGYPCGSMVTKRELLNTSEVTRPE 124
 QY 181 SSGSHSGRGFLTYASDHPDLITCLERASHYLTETESKFCPCAGCRDVAAGISGMNVGY 240
 DB 125 SSGSHSGRGFLTYASDHPDLITCLERASHYLTETESKFCPCAGCRDVAAGISGMNVGY 184
 QY 241 RPTSLCKRAIHAGIITADELGQISVLRKGISRYEGILANGVLSRDSLSDKRELFTSN 300
 DB 185 RPTSLCKRAIHAGIITADELGQISVLRKGISRYEGILANGVLSRDSLSDKRELFTSN 241
 QY 301 GGSRLSFPDQGIASSWSQSVNSGQVHMSPGQARLQDQGPWASDSSNNMKPREM 360
 DB 242 ----- 241
 QY 361 LEIDLGKKKITGRTGTSGTSGNENFYKSFVNNFNKNNKMYKGIYNNEEKVQGN 420
 DB 242 ----- 241
 QY 421 NFRDPVQNNFPIYARVRYVPTWHORIALKVELIGCQITQGNDSLWKRKTSOSTVS 480
 DB 242 ----- 241
 QY 481 TKEDETITRPISSEETSGNITTTVAIPLYLVLVFAGMGIFAFAFRKKKKSGPYSA 540
 DB 242 -----TTPGMNITVAIPSYFIALLLTGKQIFAIKRRKKKGPYSA 285
 QY 541 EAOKTDCWKQIKYPPARHQAETTSYDNEKEMTKDLTISDMA 585
 DB 286 DAOKTSCWKQIKYPPARHQAETTSYDNEKEMTKDLTISDMA 330

RESULT 2

Q9D696 PRELIMINARY; PRT; 460 AA.
 AC Q9D696
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 4631413K11RLK PROTEIN.
 GN 4631413K11RLK
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Araiwa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Araiwa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carrinci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshewski A., Yoshida K., Hasegawa Y., Kawaji H., Kontsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK014521; BAB29409.1; -
 DR MGD: MGI:1913936; 4631413K11RLK.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR004043; LCC1.
 DR Pfam: PF00431; CUB: 1.
 DR SMART: SM00042; CUB: 1.
 DR PROSITE: PS01180; CUB: 1.
 DR SEQUENCE 460 AA; 50334 MW; 3CF335670A88B93 CRC64;

Query Match
 Best Local Similarity 37.8%; Score 1154.5; DB 11; Length 460;
 Matches 237; Conservative 24; Mismatches 26; Indels 199; Gaps 1;

QY 100 MTSKNYPTVNNHYCENTITVPPKGRLLRLGLDIESQCADYLLFTSSSQGYPC 159
 DB 1 MTSKNYPTVNNHYCENTITVPPKGRLLRLGLDIESQCADYLLFTSSSQGYPC 60
 QY 160 GSMYVPEKELNTEEVVYRFESGSHISGRGFLTYASSDHPDLITCLERASHYLTETYSK 219
 DB 61 GSMYVPEKELNTEEVVYRFESGSHISGRGFLTYASSDHPDLITCLERASHYLTETYSK 120
 QY 220 FCPAGCRDVADISGNMVDGTRDTSLLCKAIIHAGIITADELGQISVLRKGISRYEGIL 279
 DB 121 FCPAGCRDVADISGNMVDGTRDTSLLCKAIIHAGIITADELGQISVLRKGISRYEGIL 180
 QY 280 ANGLVSRGSLSDKRLFTSGNCSRSLSFEPDQGIASSWSQSVNSGQVHMSPGQARL 339
 DB 181 ANGLVSRGSLSDKRLFTSGNCSRSLSFEPDQGIASSWSQSVNSGQVHMSPGQARL 198
 QY 340 QDQGPWASDSSNNMKPREMLEIDLGKKKITGRTGTSGTSGNENFYKSFVNNFNKNN 399
 DB 199 ----- 198
 QY 400 SKKRTYKGIYNNEEKVVOGNSNFPDQVQNNFPIYARVRYVPTWHORIALKVELIGC 459
 DB 199 ----- 198
 QY 460 QITQGNDSLWKRKTSOSTVSSTKKEDETTIPPISEETSGNITTTVAIPLYLVLVFA 519
 DB 199 -----TTPGMNITVAIPSYFIALLLTGKQIFAIKRRKKKGPYSA 281
 QY 520 GMGIFAFAFRKKKKSGPYSAEAKTDCWKQIKYPPARHQAETTSYDNEKEMTKDL 579
 DB 222 GMGIFAFAFRKKKKSGPYSAEAKTDCWKQIKYPPARHQAETTSYDNEKEMTKDL 281
 QY 580 ITSDMA 585
 DB 282 ITSDMA 287

RESULT 3

Q91ZV2 PRELIMINARY; PRT; 769 AA.
 AC Q91ZV2
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE ENDOTHELIAL AND SMOOTH MUSCLE CELL-DERIVED NEUROPILIN-LIKE
DE PROTEIN.
GN ESDN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX PubMed=11447234;
RA Koubek K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neutropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury";
RL J. Biol. Chem. 276:34105-34114(2001).
DR EMBL: AF387549; AAL30180.1;
SQ SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;

Query Match 30.9%; Score 942; DB 11; Length 769;
Best Local Similarity 37.5%; Pred. No. 14e-66;
Matches 230; Conservative 96; Mismatches 204; Indels 84; Gaps 19;

QY 44 PSQVMPGARGGALARA-----GRG-----LALLLAASAPRLQAEELDGGC 89
DB 16 PGRAAFAATGRAPLPAGCWCPLPGRNSSRRLLLLLLLLPDA-----GAGKGGCG 70
QY 90 HLVTYDSSGTMSTKNPYGTPYPMHTVCEKITYPKGRLLRLGLDLIE-SQTCASDYL-L 147
DB 71 HTVAGPSGTLTINPHYTPNSTCKWETRTGERIRIRFCDFIEDSDCHNYLTKI 130
QY 148 FTS---SSDQYGPYCG-SMTVPKELLNTSEVTVPFESGSHISGRGLTYASSDHPDLI 203
DB 131 ENGIGVSRTEIKYCGGLQNMOSIESGSEITVLFMSGHASGRFLASYIDKODLI 190
QY 204 TGLERASHYIKTEYSKCPAGCDVADISGNMVDGYRDTSLCKAIIHAGIADLEGQ 263
DB 191 TGLDTSVNFLEPFESKCPAGCLLPFAEISGTLPHGYRDSPLCMAGIHAGVSDVLGGQ 250
QY 264 ISVLRKGISRYEGILLANGVLSRDSGLSDKRLFTSNGSRLSFE---PGQIRASS 319
DB 251 ISVYISGTPYESSLANNTSVAGLSTLSLFEKTSQCYGLTGMESGVIAPOPTASSV 310
QY 320 WQVNESGDQVHMSPGQARLDQGPSMASGDSNNHKPREMLDLEGKKITGIRTTGS 379
DB 311 LEWTDHMGQENSMKPEKARLRKPGPPMA-FATDEH---QWLDLNLKREKKTIGIVTTGS 366
QY 380 TQSNENFYVASFYMNKNNKNNKTKYKGIYNNKEVFOGNSNFRDPVONNFIPIYARYV 439
DB 367 TLEHNTYVSAYRVLYSDQKWTYVREPGAODKIFQGNKDYHDKVDRNPLPIIARFI 426
QY 440 RVVPQTHORIALKVELIGCOIT-----QGNDSLVWRKTSQSISVSTKED 485
DB 427 RVVPQMOOKIAKVELIGCOITLKRLLKLTQPPPRKSNNL-KNTVHKLGRARF 484
QY 486 ETTRP-----IPSEETST-GINITT-----VAIPVLVLVLFAGMGFAAF- 527
DB 485 TQALQPSRNDLPRLPAQTATPDVKNLTVPFSYKDALAVLVLMALTTILILYV 544
QY 528 -----RKKKKGSPYGAESAOKTDCWKQIR--YFPAHOSAEFTISYDNE-----K 571
DB 545 CAHHMNRKKKKEGT-YDLPHWDRAGWMKGYKQLLPKASVDEETPVRYSNSSEVSLSPR 603
QY 572 EMTOKLDLITSDMA 585
DB 604 EYTTVTLQADSAEA 617

AC 0912V3.
DT 01-DEC-2001 (Tremblrel, 19, Created)
DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE ENDOTHELIAL AND SMOOTH MUSCLE CELL-DERIVED NEUROPILIN-LIKE
DE PROTEIN.
GN ESDN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR.
RX PubMed=11447234;
RA Koubek K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neutropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury";
RL J. Biol. Chem. 276:34105-34114(2001).
DR EMBL: AF387548; AAL30179.1;
SQ SEQUENCE 769 AA; 83774 MW; 73C1F164FA3F017 CRC64;

Query Match 30.7%; Score 937; DB 11; Length 769;
Best Local Similarity 37.2%; Pred. No. 3.6e-66;
Matches 229; Conservative 100; Mismatches 200; Indels 86; Gaps 17;

QY 44 PSQVMPGARGGALARA-----GRG-----LALLLAASAPRLQAEELDGGC 91
DB 16 PGRAAFAATGRAPLPAGCWCPLPGRNSSRRLLLLLLLL-LQDAGGCGGCGCHT 72
QY 92 VTYDSSGTMSTKNPYGTPYPMHTVCEKITYPKGRLLRLGLDLIE-SQTCASDYL-L 149
DB 73 VLPESGTLTINPHYTPNSTCKWETRTGERIRIRFCDFIEDSDCHNYLTKIFN 132
QY 150 S---SSDQYGPYCG-SMTVPKELLNTSEVTVPFESGSHISGRGLTYASSDHPDLITC 205
DB 133 GIGVSRTEIKYCGGLQNMOSIESGSEVTLFMSGHASGRFLASYIDKEDLITC 192
QY 206 LERASHYIKTEYSKCPAGCDVADISGNMVDGYRDTSLCKAIIHAGIADLEGQIS 265
DB 193 LDTVSNFLEPFESKCPAGCLLPFAEISGTLPHGYRDSPLCMAGIHAGVSNVLGGQIS 252
QY 266 VLQRKISRREGLANGVLSRDSGLSDKRLFTSNKCSRLSFE---PDQIRASSWQ 321
DB 253 TVLSKGTPIYESSLANNTSVAGLSTLSLFEKTSQCYGLTGMESGVIAPOPTASSALE 312
QY 322 SVNESGDQVHMSPGQARLDQGPSMASGDSNNHKPREMLDLEGKKITGIRTTSTQ 381
DB 313 WTDHMGQENSMKPEKARLRKPGPPMA-FATDEH---QWLDLNLKREKKTIGIVTTGS 368
QY 382 SNENFYVASFYMNKNNKNNKTKYKGIYNNKEVFOGNSNFRDPVONNFIPIYARYV 441
DB 369 IEHSYVSAVRYLYSDQKWTYVREPGVDOKIFQGNKDYHDKVDRNPLPIIARFI 428
QY 442 VPQTHORIALKVELIGCOIT-----QGNDSLVWRKTSQSISVSTKED 474
DB 429 NPVQMOOKIAKVELIGCOITLKRLLKLTQPPPRKSNNL-LKNTVHKLGRARF 485
QY 475 -----OSTVSTRKEDEETTRPIPEETSTGINITVVAIPVLVLVLFAGMGFAAF- 527
DB 486 QVLOPSRNDLPRLPAQTATPDVKNLTVPFSYKDALAVLVLMALTTILILYV 545
QY 528 -----RKKKKGSPYGAESAOKTDCWKQIRPF-----ARHOSAEFTISYDNE 570
DB 546 AMHMRNRKKKKEGA-YDLPHWDRAGWMKGYKQLLPKASVDEETPVRYSTSE--VSHLSA 602
QY 571 KEMTOKLDLITSDMA 585
DB 603 REVYTVLQADSAEA 617

RESULT 5
ID Q96PD2 PRELIMINARY; PRT: 775 AA.
AC Q96PD2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE ENDOTHELIAL AND SMOOTH MUSCLE CELL-DERIVED NEUROPILIN-LIKE PROTEIN.
GN ESDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RX SEQUENCE FROM N.A.
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the Longest Secretary Signal Sequence among Eukaryotes, Is up-regulated after Vascular Injury";
RL J. Biol. Chem. 276:34105-34114(2001).
DR EMBL; AF387547; MAF30178.1; -.
SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Query Match 30.4%; Score 929; DB 4; Length 775;
Best Local Similarity 36.3%; Pred. No. 1.6e-65;
Matches 223; Conservative 106; Mismatches 205; Indels 84; Gaps 17;

QY 36 EAEIAPLPGVAVPGARGGALARAARGL--LALLAVASAPRLQAELGDCGHLV 92
DB 17 QVRAAAAPAPMALPLSRSLPCSSNSSFSMPFLLLVLLLEDAQAQDCGCHV 76
QY 93 TYDQSGMTSKNPTGYPNHTVEKITYPKGRLLRLRGDLDIE-SQTCASDIYLFIS- 150
DB 77 LGPSSGLTINYPQYPNSTVCWEIRKMGERVRIKGDIDEDSOSCHFNRYING-136
QY 151 ---SSDQYPGYCG-SMTVPKELLNTSEVTVRESGSHISGRGFLITVYSSPHPLITCL- 206
DB 137 IGSRTPIGKTCGLGLQMHNSISKGNETITLFMSGIHVSGRGLASTYVIDKODITCL 196
QY 207 ERASHYLTKEYSKFCPCAGCDVAGDISGMNVGDRDTSLLCKRAAIIAGIADLGQISV 266
DB 197 DTSNFELEPESKYPACGLLPFAEISGITIPHCYRDSPLCMAGVAGVSNLTGQISV 256
QY 267 LQKRGISRYEGILANGVLSHDSLSDKRFLTNSGCSRLSFE---PDGQIRASSMS 322
DB 257 VISGIDIPYESSLANNVTSVGHLSLTFPKTSGCYGLGMEGVIADPQITASSVLEW 316
QY 323 VNESGDVHWSPGOARLQDQPSMASGDSNNHKNPREMLEIDLGEKKITGTRTGSTOS 382
DB 317 TDHTGOENSKPKKARLAKKGPMAAFATDE---YQWQIDLNKEKKTITGITGSTIV 372
QY 383 NFNFYVSEFWNFKNNNSKKYKGIYNNDEKVFQSGNSFROPVONNIPPIVARYRV 442
DB 373 EHNYYSAVRILYSDQKWTYREPGVEODKIFQGNKMYHODVRRNFPPIIARFIRVN 432
QY 443 POTHORIALKVELIGQ-----ITQ-----GNSIWRKTSOSTVSTKEDEBT 487
DB 433 PTQWQKTIAMKMLGQIFPKGRPKLTQPPPRNSND---LKNYTAAPPIAIGRAPK 488
QY 488 ITRPI-----PSEET--STGINITT-----VAIPVLVLVLVAFAGMGIFAFA 527
DB 489 FTQPLQPRSSNEPPAQVETASPDIRNTYVFNPKDVALAALVPLVLMVLTLLIL 548
QY 528 -----RKKKKGS---PYGSAEAKQTDCKQIKYPP-----ARHQAETPI 565
DB 549 VCAHMHNRKRRKTEGTIDLPYW---DRAGWMKGMKQFLPAKAVDHEETPVRYSSSE-V 602
QY 566 SYDNKEMTQKIDLTISDMA 585

DB 603 NHLSPREVTTVLQADSAEYA 622
RESULT 6
ID Q14089 PRELIMINARY; PRT: 364 AA.
AC Q14089;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HYPOTHETICAL 40.0 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RX SEQUENCE FROM N.A.
RX Shludate T.;
RA "unpublished";
RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; D29810; BAA18909.1; -.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR004042; FA58_C.
DR InterPro; IPR004043; LCCL.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Query Match 18.8%; Score 572.5; DB 4; Length 364;
Best Local Similarity 38.6%; Pred. No. 1.6e-37;
Matches 147; Conservative 61; Mismatches 124; Indels 49; Gaps 16;

QY 86 DGGCHLVTYDQSGMTSKNPTGYPNHTVEKITYPKGRLLRLRGDLDIE-SQTCASD 144
DB 1 DCGCHTVLPGESGTLINYPQYPNSTVCWEIRKMGERVRIKGFDFIEDSDSCHFN 60
QY 145 YLLEFS-----SSDQYPGYCG-SMTVPKELLNTSEVTVRESGSHISGRGFLITVYSSPH 199
DB 61 YLRVYNGISVREIKYCGGLQMHNSISKGNETITLFMSGIHVSGRGLASTYVID- 119
QY 200 PDLITCLERASHYLTKEYSKF-----CPAGCDVAGDISGMNVGDRDTSLLCKRAA 250
DB 120 -----KQSNLTFEGHCIOFPNNLSYSTAPACGLPFAEISGITIPHCYRDSPLCMAG 172
QY 251 IHAGIADLGQISVYLQKRGISRYEGILANGVLSHDSLSDKRFLTNSGCSRLSFE 309
DB 173 VHAGVSNLTGQISVYISKGIYESSLANNVTSVGHLLIQVEFFKTSQCYOTLQME 232
QY 310 PDQIRASS-----WQSVNESGDVHWSPGOARLQDQPSW-ASGDSNNHKNPREML 361
DB 233 SGGD-RGSSNNSITVLEW--TDHTGOENSKPKKSAE---KTWALGAFATDE--YQWL 284
QY 362 EIDGEEKKKTGTRTGSTQ-SNFNFYVKS---FWNFKNNNSKKYKGIYNNDEKVFQ 417
DB 285 QIDLNKEKKTITGITGSTIVSITIMCLPTESQTYVMGKNG---LCRESLVEQDKIFQ 340
QY 418 GNSN-FRDPVONNIPPIYAR 437
DB 341 GNRKILTRVVRNNELPIIAR 361
RESULT 7
Q9H2E2
ID Q9H2E2 PRELIMINARY; PRT: 555 AA.
AC Q9H2E2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
GN NEUROFILIN-2 SOLUBLE ISOFORM 9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATIC ADENOCARCINOMA;
RX MEDLINE-20564205; PubMed-11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT *Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.*;
RL Genomics 70:211-222(2000).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR EMBL; AF280546; AAG41403.1; .
DR HSSP; P12259; IC2T.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00431; CUB; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C; 1; FA58C_1.
SQ SEQUENCE 555 AA; 62430 MW; B67B7C992AB0546F CRC64;

Query Match 11.4%; Score 349; DB 4; Length 555;
Best Local Similarity 25.3%; Pred. No. 2.3e-19;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

QY 88 CGHLVYQDSGTMSTKPYGTYPNHTVCEKITYPK-GKRLILRLG-DLDIESQTCASDY 145
DB 28 CGRLNSDKAGYITSPGTPDPSHONCEWIVYAEPNOKIVLNFNPHFIEKHCKYDF 87
QY 146 LLF-----TSSDOYGPYCGSMTPVKELLNTSEVTVRESGSHISGRGFLTY-----AS 196
DB 88 IEIRDGDESADLIGKHGCI-APPTLISSGSMYIKFTSDYAROGAGFSLRYEFKGS 146
QY 197 SD-----HPDLITC-----LERASHYKTEYSKF----- 220
DB 147 EDCSKNTSPNGTIESPGFPEKYPHNLDCTFTILAKPKMETILQLFLFDLEHDPLOVGE 206
QY 221 -CPAGCRDVAGDIS--GNMVDGYRDTSLCKAAIHAGIADLGQISVLOKGISRYEG 277
DB 207 DCKYDMDIDWDGIPHVGPRLKYGCTKTPSELSSSTGLSLFHTDMAVAK-----DG 259
QY 278 ILANGVLSRSGSLDKRFLFTSNCGSRSLSF-----PDQIRASSWQSVNESGDQVHS 333
DB 260 FSARYLVHGEPLNFQ-----CNVPLGMEGRIANEOISASSTY-----SDGRWT 305
QY 334 PGQARLDQDGPWASGDSNNHKKPREMLEIDLGKKKITGIRTTG--STOSNENFYVKSF 391
DB 306 PQGSRHLGDNQGTWPNLDSN---KEYLOYDLRFLMTALATOGAISRETONGYIVKSY 361
QY 392 VAMFNKNNSKMKTYYKGIIVNEEKVFGNSNFRDPVONNFPPIYARVYRVVQTHORIA 451
DB 362 KLEVSTNGEDMWMYVRHGNH--KVFQANNDATEVYLKRLHAPLILTRVRIIPQTHSGIA 419
QY 452 LKVELIGCOIT 462
DB 420 LRLELFGCRVY 430

RESULT 8
ID Q9H2E4 PRELIMINARY; PRT; 901 AA.
AC Q9H2E4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE NEUROFILIN-2B(O).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20564205; PubMed-11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT *Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.*;
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF280544; AAG41403.1; .
DR HSSP; P12259; IC2T.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; FA58C_1.
DR PROSITE; PS50060; MAM_2; 1.
DR GlycoProtex; Neuropeptide Receptor; Transmembrane.
SQ SEQUENCE 901 AA; 101380 MW; 1F3107A9CC665EA2 CRC64;

Query Match 11.4%; Score 349; DB 4; Length 901;
Best Local Similarity 25.3%; Pred. No. 4.8e-19;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

QY 88 CGHLVYQDSGTMSTKPYGTYPNHTVCEKITYPK-GKRLILRLG-DLDIESQTCASDY 145
DB 28 CGRLNSDKAGYITSPGTPDPSHONCEWIVYAEPNOKIVLNFNPHFIEKHCKYDF 87
QY 146 LLF-----TSSDOYGPYCGSMTPVKELLNTSEVTVRESGSHISGRGFLTY-----AS 196
DB 88 IEIRDGDESADLIGKHGCI-APPTLISSGSMYIKFTSDYAROGAGFSLRYEFKGS 146
QY 197 SD-----HPDLITC-----LERASHYKTEYSKF----- 220
DB 147 EDCSKNTSPNGTIESPGFPEKYPHNLDCTFTILAKPKMETILQLFLFDLEHDPLOVGE 206
QY 221 -CPAGCRDVAGDIS--GNMVDGYRDTSLCKAAIHAGIADLGQISVLOKGISRYEG 277
DB 207 DCKYDMDIDWDGIPHVGPRLKYGCTKTPSELSSSTGLSLFHTDMAVAK-----DG 259
QY 278 ILANGVLSRSGSLDKRFLFTSNCGSRSLSF-----PDQIRASSWQSVNESGDQVHS 333
DB 260 FSARYLVHGEPLNFQ-----CNVPLGMEGRIANEOISASSTY-----SDGRWT 305
QY 334 PGQARLDQDGPWASGDSNNHKKPREMLEIDLGKKKITGIRTTG--STOSNENFYVKSF 391
DB 306 PQGSRHLGDNQGTWPNLDSN---KEYLOYDLRFLMTALATOGAISRETONGYIVKSY 361
QY 392 VAMFNKNNSKMKTYYKGIIVNEEKVFGNSNFRDPVONNFPPIYARVYRVVQTHORIA 451
DB 362 KLEVSTNGEDMWMYVRHGNH--KVFQANNDATEVYLKRLHAPLILTRVRIIPQTHSGIA 419
QY 452 LKVELIGCOIT 462
DB 420 LRLELFGCRVY 430

```

RESULT 9
ID Q9H2D5 PRELIMINARY; PRT; 901 AA.
AC Q9H2D5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEUROFILIN-2B(0).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms."
RT Isoforms."
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF281074; AAC41899.1; -.
DR HSSP: P12259; 1CZP.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR000998; MAM.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1;
DR PROSITE: PS50060; MAM; 2; 1.
KW Glycoprotein; Neurone; Receptor; Transmembrane.
SQ SEQUENCE 901 AA; 101381 MW; A1310304AFD086AB CRC64;

Query Match 11.4%; Score 349; DB 4; Length 901;
Best Local Similarity 25.3%; Pred. No. 4.8e-19;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

QY 88 CGHLVYQDSGVTMSKNTPTVPHNTVCEKTTVPR-KRLLRLRLG-DLDTESQTCASDY 145
DB 28 CGRLNSKDAGYITSPGYPODYPSHONCEVIYAPENCKIVLNNPHELEKHKDCKYDF 87
QY 146 LLF-----TSSDQGYPCGSMVPRKELLNTSEVTRFESGSHISGRFLLTY-----AS 196
DB 88 IEIRDCSEADLLGKHCNIT-APPTIISGSMLYIKFTSDYARAGAGSLRYELFKGS 146
QY 197 SD-----HPDLITC-----LERASHYLKTEYSKF----- 220
DB 147 EDCSKNFTSPNGTIESPPEKYPHNLDCFTFLAKPKMEITLLQFLIFLEHDPLOYEG 206
QY 221 -CPAGCDVAGDIS--GNMYDGYRDTSLCKAIAIHAGITADELGGQISVLQKKGSRTEG 277
DB 207 DCKDWMIDWDGIPHVGPPLGKYGCTKPTSELSSYGLISLTFPHDMAVAK-----DG 259
QY 278 ILANGVYSRSGISLDRFLFTSNGCSRSLFE-----PDGQIRASSSQWSVNSGDOVHMS 333
DB 260 FSARYTYVHQEPLNEFQ-----CNVPLGMEGRIANQISASSY-----SDGRYT 305
QY 334 PGARLLDQGPASVAGSDSNHNRPRMELTIDLEKKKIKGITRTG--STOSNFNFYVSF 391
DB 306 PQGSRLAGDNGRTPNLDSN-----KEYLOYDLFLMLTALATQGISLSETONGYIVKSY 361
QY 392 YVNEKNNNSKWKTYKGIIVNEKEKVFQGSNSFRDPVONNFPPIVARYVAVDQTHORIA 451

```

```

DB 362 KLEVSTNGEDMNVYHRGKNH--KVFQANNDATEVYLNKHAFLTRFVIRIPQTHWSGIA 419
QY 452 LKVELIGCOIT 462
DB 420 LRLEFGCRVT 430

RESULT 10
ID Q9H2E3 PRELIMINARY; PRT; 906 AA.
AC Q9H2E3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEUROFILIN-2B(5).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms."
RT Isoforms."
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF280545; AAC41404.1; -.
DR HSSP: P12259; 1CZP.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR000998; MAM.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1;
DR PROSITE: PS50060; MAM; 2; 1.
KW Glycoprotein; Neurone; Receptor; Transmembrane.
SQ SEQUENCE 906 AA; 101955 MW; 0FB01EF24D8D9B0BD CRC64;

Query Match 11.4%; Score 349; DB 4; Length 906;
Best Local Similarity 25.3%; Pred. No. 4.9e-19;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

QY 88 CGHLVYQDSGVTMSKNTPTVPHNTVCEKTTVPR-KRLLRLRLG-DLDTESQTCASDY 145
DB 28 CGRLNSKDAGYITSPGYPODYPSHONCEVIYAPENCKIVLNNPHELEKHKDCKYDF 87
QY 146 LLF-----TSSDQGYPCGSMVPRKELLNTSEVTRFESGSHISGRFLLTY-----AS 196
DB 88 IEIRDCSEADLLGKHCNIT-APPTIISGSMLYIKFTSDYARAGAGSLRYELFKGS 146
QY 197 SD-----HPDLITC-----LERASHYLKTEYSKF----- 220
DB 147 EDCSKNFTSPNGTIESPPEKYPHNLDCFTFLAKPKMEITLLQFLIFLEHDPLOYEG 206
QY 221 -CPAGCDVAGDIS--GNMYDGYRDTSLCKAIAIHAGITADELGGQISVLQKKGSRTEG 277
DB 207 DCKDWMIDWDGIPHVGPPLGKYGCTKPTSELSSYGLISLTFPHDMAVAK-----DG 259
QY 278 ILANGVYSRSGISLDRFLFTSNGCSRSLFE-----PDGQIRASSSQWSVNSGDOVHMS 333

```



```
Db 260 FSARYLVHQPENFO-----CNVPLGMEGRIANEQISASTY-----SDGRWT 305
OY 334 PGQARLDQGPWASGSSNNHKKRELELDGKKKITGRTG--STQSNFNFYVKSF 391
Db 306 PQOSRLHGDDNGWTPNIDSN---KEYLOYDLFELTMTAIOGAISETONGYVYKSY 361
OY 392 VAMFNKNNKSKKTYKGIIVNNEEKVFGNSNFRDPYONNFIPIVARYVVPOTWHQRTA 451
Db 362 KLEVSTNGEDMVMYRHQKNH--KVFOANNDATEVVLNKLHAPLTLRFVIRPQTWHSGIA 419
OY 452 LKVELLIGCQTT 462
Db 420 LRLELFCRCVY 430

RESULT 11
ID 09H2D4 PRELIMINARY; PRT; 906 AA.
AC 09H2D4;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE NEUROPIILIN-2B(5).
OS NRP2.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms."
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPIILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF281074; AAC41900.1; -.
DR HSSP; P12259; ICDT.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01185; FA58C_1; UNKNOWN_1.
DR PROSITE; PS50060; MAM_2; 1.
KW Glycoprotein; Neurone; Receptor; Transmembrane.
SQ SEQUENCE 906 AA; 101956 MW; EB183F265457B0B9 CRC64;

Query Match 11.4%; Score 349; DB 4; Length 906;
Best Local Similarity 25.3%; Pred. No. 4.9e-19;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

OY 88 CCHLVYODSGTMTSKNPGYTPNITVEKTIYVK-GKRLILRG-DLDIESQTCASDY 145
Db 28 CGRLNSKADAGITTPGTPDIPSHQNCWEIVLAPPEPQOKVLNPNHFEIEKHDKYDF 87
OY 146 LTF-----TSSDQGYPCYSMTVPKELLNLTSEVTVRESGSHISGRFLLTY-----AS 196
Db 88 IETRDGDSADLGLKHCNGNI-APPTLISSGSMLYIKFTSYANQAGGFSLRITFKTGS 146
OY 197 SD-----HPDLITC-----LERASHILKTEYSKF----- 220
Db 147 EDCSKNFTSPNGTIESPGFPEKYPHNLDCTPTILAKPKMEIILLOFLFLEHDPLOYGEG 206
```

```
OY 221 -CPACRVDVAGDIS--GNMVDGYRDTSLCKAALHAGIINDELGOISYLQKGISREYEG 277
Db 207 DCKTQMDIDMGIRHVPGLGKTCGTTPSELSSGTIISLTFPHDMAVK-----DG 259
OY 278 ILANGVLSRDGSLSDKRFLLTSSNGCSRLSFE---PDQIRASSWSQSVNESGQVHMS 333
Db 260 FSARYLVHQPENFO-----CNVPLGMEGRIANEQISASTY-----SDGRWT 305
OY 334 PGQARLDQGPWASGSSNNHKKRELELDGKKKITGRTG--STQSNFNFYVKSF 391
Db 306 PQOSRLHGDDNGWTPNIDSN---KEYLOYDLFELTMTAIOGAISETONGYVYKSY 361
OY 392 VAMFNKNNKSKKTYKGIIVNNEEKVFGNSNFRDPYONNFIPIVARYVVPOTWHQRTA 451
Db 362 KLEVSTNGEDMVMYRHQKNH--KVFOANNDATEVVLNKLHAPLTLRFVIRPQTWHSGIA 419
OY 452 LKVELLIGCQTT 462
Db 420 LRLELFCRCVY 430

RESULT 12
ID 096I90 PRELIMINARY; PRT; 609 AA.
AC 096I90;
DT 01-DEC-2001 (Tremblrel, 19, Created)
DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE SIMILAR TO NEUROPIILIN 1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007737; AAH07737.1; -.
SQ SEQUENCE 609 AA; 68347 MW; 798AAMD2D568C127 CRC64;

Query Match 11.0%; Score 335; DB 4; Length 609;
Best Local Similarity 24.9%; Pred. No. 3.5e-18;
Matches 124; Conservative 76; Mismatches 176; Indels 122; Gaps 23;

OY 64 RGL-----LALLAVSAPRLQAELELDGCGHLVYODSGTMTSKNPGYTPNITVEK 117
Db 3 RGLPLCAVLALVLPAGAFR-----NDKCGDITKIESPGLYLPSPGPHSHSEKCEW 56
OY 118 TTVPRKG-KRLILRG-DLDIESQTCADYL-LTSSSDQ---YGPQSGSKTVPKELLN 171
Db 57 LQADPPYQYRIMININPNHFDLEDRCCKYDVEVDGENENHGFGRKCGKI-APPVYSS 115
OY 172 TSEVTVRESGSHISGRFLLTYA-----SSDHPDLITC-- 205
Db 116 GPELFIFKFSVDYETHGAGSIRYEIFKRGPECSQNTTTPSPVIKSPGFPEKIPNSLECTY 175
OY 206 -----LERASHILKTEYSKFPGAG--CDVAGDISGNMVDGYRDTSLCKRAIH 252
Db 176 IVFAPRMESEILIEFESFLEPDSNP--PGGMFCRYDRLEI---WDGFPD-----VGP 223
OY 253 AG-TIADELGGQI-----SVLQKGISREGLIANGVLSRGSISDKRFLF 297
Db 224 IGRYCGQKTPGRIRSSSILSMVFTYTSALAKBEFSANYSVLQSSV-SEDFK----- 274
OY 296 TSNCGSRSLSEP---DQGITRASSWSQSVNESGQVHMSPGQARLDQGPWASGSSN 353
Db 275 ---CMAIGMESGRFIHSDQITASSQYST-----NMAEASRLNYPNGTTPBEDSY 322
OY 354 NHPREMLELDGKKKITGRTG--STQSNFNFYKSVVAMFNKNNKSKKTYKGIIVN 411
```

Db 323 ----REMIQVLDGLLEFVAVAGTQGAISKEFKKKYVYTKYKIDVSSNGEDMTIK--EGN 376
OY 412 EEKVEGCGNSFRDPVONNFIPIVARYVAVPQTHQRIALKEVLIGQIT---QGDS 467
Db 377 KPYLFGGNTNPPDVVAVVAPKPLITRFVRIKATWETGISMRFEYGGKIDYPCSGMLG 436
OY 468 LVMKTSOSTSVSTKED 485
Db 437 MWSGLISDSQITSSNOGD 454

RESULT 13

O96IH5 PRELIMINARY: PRT: 644 AA.
AC O96IH5
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEUROPIILIN 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY, AND RENAL CELL ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC007533; AAH07533.1; -
SQ SEQUENCE 644 AA; 71907 MW; 9204B6F5C3C2A21 CRC64;

Query Match 11.0%; Score 335; DB 4; Length 644;
Best Local Similarity 24.9%; Pred. No. 3.8e-18;
Matches 124; Conservative 76; Mismatches 176; Indels 122; Gaps 23;

OY 64 RGL-----LALLAVSAPRLQAEELDGCGHLVYTDGSGTMTSKNPGTYPNHVCER 117
Db 3 RGLPLCAVALVALVAPAGAFR-----NDKCGDTIKIESPGYLTSPGPHSHYHSEKCEW 56
OY 118 TIVPFGK-KRLILRLG-DDDISQTCASDYL-LFTSSSDQ---YGPYCGSMVFPRELLN 171
Db 57 LIOAPPYQRIIMINFNPHDLEDROCKYDYEVFDGENENHFRKFCGKI-APPVYSS 115
OY 172 TSEVTVRFESGSHISGRGFLTYA-----SSDHPDLITC-- 205
Db 116 GPFLFKIFYSDYETHGAGSIRYEIFKRGPECSQNTTPSGVIKSPGPEKYPNSLECTY 175
OY 206 -----LEASHYIKTEYSKFCPAG--CRDVADDISGMNVDYGRDTSLLCKAAIH 252
Db 176 IYFAPRMSEILIEFESFDLEPDSNP--PGMFCRYDRLEI---WDGFPD-----VGPH 223
OY 253 AG-TIADLEGQI-----SVLQKGISRYEGILANGVLSRDGSLSDKRFLE 297
Db 224 IGRYCGQKTPGRIRSSSGILSMVFYTDASIAKEGFSANYSVLQSSV--SEDPK----- 274
OY 298 TSNCGCRSLSFEP---DQOIRASSWQSVNMSGQVHMSRQALDQOGGSMASGSSN 353
Db 275 ----CHEALGMSGEIHSQITASSQYSF-----NMSAERSRLNYPENGWTPGEDSY 322
OY 354 NHRPREMLEIDGEEKKIKIGIRITG--STQSNFNFYVKSFWANFNKNNNSKKYTKGIYNN 411
Db 323 ----REMIQVLDGLLEFVAVAGTQGAISKEFKKKYVYTKYKIDVSSNGEDMTIK--EGN 376
OY 412 EEKVEGCGNSFRDPVONNFIPIVARYVAVPQTHQRIALKEVLIGQIT---QGDS 467
Db 377 KPYLFGGNTNPPDVVAVVAPKPLITRFVRIKATWETGISMRFEYGGKIDYPCSGMLG 436
OY 468 LVMKTSOSTSVSTKED 485
Db 437 MWSGLISDSQITSSNOGD 454

RESULT 14

O9H2E1 PRELIMINARY: PRT: 704 AA.
AC O9H2E1
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEUROPIILIN-1 SOLUBLE ISOFORM 11.
GN NRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE ADENOCARCINOMA;
RX MEDLINE-20564205; PubMed-11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neupilin-1 and Neupilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.";
RL Genomics 70:211-222(2000).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR EMBL: AF280547; AAC41406.1; --
DR HSSP: P12259; IC2F.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00431; CUB; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; 1.
DR PROSITE: PS01285; FA58C_1; 1.
SQ SEQUENCE 704 AA; 79002 MW; 1625DD561459307 CRC64;

Query Match 11.0%; Score 335; DB 4; Length 704;
Best Local Similarity 24.9%; Pred. No. 4.4e-18;
Matches 124; Conservative 76; Mismatches 176; Indels 122; Gaps 23;

OY 64 RGL-----LALLAVSAPRLQAEELDGCGHLVYTDGSGTMTSKNPGTYPNHVCER 117
Db 3 RGLPLCAVALVALVAPAGAFR-----NDKCGDTIKIESPGYLTSPGPHSHYHSEKCEW 56
OY 118 TIVPFGK-KRLILRLG-DDDISQTCASDYL-LFTSSSDQ---YGPYCGSMVFPRELLN 171
Db 57 LIOAPPYQRIIMINFNPHDLEDROCKYDYEVFDGENENHFRKFCGKI-APPVYSS 115
OY 172 TSEVTVRFESGSHISGRGFLTYA-----SSDHPDLITC-- 205
Db 116 GPFLFKIFYSDYETHGAGSIRYEIFKRGPECSQNTTPSGVIKSPGPEKYPNSLECTY 175
OY 206 -----LEASHYIKTEYSKFCPAG--CRDVADDISGMNVDYGRDTSLLCKAAIH 252
Db 176 IYFAPRMSEILIEFESFDLEPDSNP--PGMFCRYDRLEI---WDGFPD-----VGPH 223
OY 253 AG-TIADLEGQI-----SVLQKGISRYEGILANGVLSRDGSLSDKRFLE 297
Db 224 IGRYCGQKTPGRIRSSSGILSMVFYTDASIAKEGFSANYSVLQSSV--SEDPK----- 274
OY 298 TSNCGCRSLSFEP---DQOIRASSWQSVNMSGQVHMSRQALDQOGGSMASGSSN 353
Db 275 ----CHEALGMSGEIHSQITASSQYSF-----NMSAERSRLNYPENGWTPGEDSY 322
OY 354 NHRPREMLEIDGEEKKIKIGIRITG--STQSNFNFYVKSFWANFNKNNNSKKYTKGIYNN 411
Db 323 ----REMIQVLDGLLEFVAVAGTQGAISKEFKKKYVYTKYKIDVSSNGEDMTIK--EGN 376
OY 412 EEKVEGCGNSFRDPVONNFIPIVARYVAVPQTHQRIALKEVLIGQIT---QGDS 467
Db 377 KPYLFGGNTNPPDVVAVVAPKPLITRFVRIKATWETGISMRFEYGGKIDYPCSGMLG 436
OY 468 LVMKTSOSTSVSTKED 485
Db 437 MWSGLISDSQITSSNOGD 454

Db 437 MVSGLISDSQITSSNOCD 454

RESULT 15

090X38 PRELIMINARY; PRT; 921 AA.

ID 090X38

AC 090X38;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE NEUROPLIN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_Taxid=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97433084; PubMed-9288753;

RA He Z., Tessier-Lavigne M.;

RT "Neuropilin is a receptor for the axonal chemorepellent Semaphorin III.";

RT Cell 90:739-751(1997).

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE NEUROPLIN FAMILY.

CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

DR EMBL: AF018957; AAC53345.1; -.

DR HSP; P12259; IC2P.

DR InterPro: IPR000859; CUB.

DR InterPro: IPR000421; FA58_C.

DR InterPro: IPR001092; HLR_dlm.

DR InterPro: IPR000998; MAM.

DR Pfam: PF00431; CUB; 2.

DR Pfam: PF00754; FS_F8_type_C; 2.

DR Pfam: PF00629; MAM; 1.

DR PRINTS: PR00020; MAMDOMAIN.

DR SMART: SM00231; FA58C; 2.

DR SMART: SM00137; MAM; 1.

DR PROSITE: PS01180; CUB; 2.

DR PROSITE: PS01285; FA58C_1; 2.

DR PROSITE: PS01286; FA58C_2; 2.

DR PROSITE: PS00740; MAM_1; 1.

DR PROSITE: PS50060; MAM_2; 1.

KW Glycoprotein; Neurone; Receptor; Transmembrane

SEQUENCE 921 AA; 103052 MW; 58B29A9AA4978971 CRC64;

Query Match 10.7%; Score 325.5; DB 11; Length 921;

Best Local Similarity 24.0%; Pred No. 3.8e-17;

Matches 110; Conservative 81; Mismatches 173; Indels 95; Gaps 18;

QY 64 RGLALLAVSAPRLQAEELDGCGLVYQDSGTMTSKNYPGYVPHNTVCEKTIYVK 123

Db 3 RGLPILCATIALAIALGAFR-SDKCGGFIKINPGYILSPGYHSEKCEMLIQAPE 61

QY 124 G-KRLIRLG-DLDIESQTCADYLLFTSSDQ---YCPYCGSMIVPKELLINTSEVY 177

Db 62 PYORIMINFNPHDLDEDRCKDYVEVIDGENEGRLMGKFCGKI-APSPVYSSGFLEFI 120

QY 178 RFESGSHISGRGFLTYA-----SDHPDLITC----- 205

Db 121 KFYSDVETHGAGPSIRIEIFKRGPECSQNYTAPTVIKSPGFPEKYPNSELECTYIIIFAPK 180

QY 206 -----LERASHYLKTEYSKFCPAG--CRDVAGDISGMVVDGYRDTSLCKAAIHAGIAD 258

Db 181 MSEIILIEFESFLEQDSNP--PGMFCGRYDRLEI---WDGPE-----VGP 221

QY 259 ELGGQISVLRKGIKSYEGIIA-----NGVLSRDSLSDKRFLFTSNG-----CSRSLSE 309

Db 222 HIGRYCGQKTPGRIRSSGILSMVFYDTSAIAKEGFSNANYSVLQSSISSEDFKMEALGME 281

QY 310 P-----DGIIRASSSWQSYNESGDQYHWSPGQARLQDGPWSAGSDSSNNHPRMELIDL 365

Db 282 SCEIHSDOITASSQYGT-----NMSVRSRLNYPENGWTPGEDSY----REMIQVDL 329

QY 366 GEKKKITGIRTTG--STQSNFNFYKSFVMMFKNNNSKWKTYKGIYVNNEEKVPGQNSNR 423

Db 330 GILRFVTAVGTQGAISKETKKRYKYRYVDISSNGEDMTLK--EGNKALIFQGNINPT 387

QY 424 DPVQNNFIPPIYARYRVVQPMHORIALKVELIGCOIT 462

Db 388 DYVEGVFPKPLITRFVRIRKIPASMETGISMRFEVYCKIT 426

Search completed: July 9, 2002, 02:25:21

Job time: 678 sec

Query Match 100.0%; Score 1761; DB 6; Length 1761;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggatcgcgtgagggcagacgacgccccgtcccgccgcccgcgtgctgcgaagag 60
DB 1 ATGGATTCGGTGC GGCGGACGACCTGGCCCTCCGGCGCGCTCTCCGAGAG 60
QY 61 gagggcccccggcagcgtcgagctcggaaccgtcgaaggagggcgaagctgccc 120
DB 61 GAGGGCGCCCGCGGCGGACCTGGGCTCGGATCCGTGAGGGAGGCGGAGCTTCCC 120
QY 121 aagctggcgcccgagcggggcactggtccggcgcccgaggcgagcgcaatggcg 180
DB 121 AAGCTGGCGCCCGGCGGCTCATGTGCCCGCGCGCGCGCGGCGGCGCACTGGCGCGG 180
QY 181 gctgcggggcgggcctcctgctgctgctgctgctgctgctgctgctgctgctg 240
DB 181 GCTGCCGGCGGGGCTCTCTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 241 gcggagagcctggtgctgagctgctgagacactgactgactgactgactgactg 300
DB 241 GCGGAGAGACTGGTGTGATGGCTGTGACACTAGTACTATGAGTGTGAGTGTGAGTGTG 300
QY 301 aactcgaagaattatcccgaggacccaactcaactgctgctgcaaaagaactaca 360
DB 301 ACATCTAAGAAATTATCCGGGACCTACCCCAATCACCTGTTGGGAAAGACAAATTACA 360
QY 361 gtaccaaagggaagaaactgactgaggttgagagattggaatcccaagacc 420
DB 361 GTACCAAGAGGAAAGAACTGATCTGAGGTTGGGAGATTGGATATGAAATCCAGACC 420
QY 421 tggctcctgactatcctcctcctcctcctcctcctcctcctcctcctcctcct 480
DB 421 TGGCTCTTGACTATCTTCTTCTTCCAGCTTTCAGATCAATATGTCCTACTGTGGA 480
QY 481 agatgactgctcccaaaagaactctgttgaacacaaagtgaagtaagctcgcctt 540
DB 481 AGATGACTGCTCCCAAAAGAACTCTTGTGAAACAAAGTGAAGTAACTCCGCTTGAG 540
QY 541 agtgaatcccaactctcggcggggttcttgcctgacctgacgagcagacatcca 600
DB 541 AGTGAATCCCAACTCTTCTGCGCGGGTCTTGTGACTATGACGAGCAGACATCCA 600
QY 601 gatttaatacatgcttgaagagcgtgacatacttgaagacgaataaagaataatc 660
DB 601 GATTTAATAACATGTTTGAAGAGCTACGCCATTAATTTGAAGACGAATTAACGAATTC 660
QY 661 tggccagcctggtgtgagagacgtaagaagaacatcttcggaataatgtagatgat 720
DB 661 TGGCCAGCTGCTGTGAGAACCTAGCAGAGACATTTCTGGGAATAATGTGATGATGAT 720
QY 721 agagatacctccttattgtgcaaaagctgcacatcagagaaataatgctgataacta 780
DB 721 AGAGATACCTCTTATTGTGCAAAAGCTGCCATCATGAGGAATTAATGCTGATGAATTA 780
QY 781 ggtggccagacatcgtgctgctgctgctgctgctgctgctgctgctgctgctg 840
DB 781 GGTGGCCAGATCATGCTGCTTCAAGCCAAAGGATCATGCTATTAAGGAGATTTGGCC 840
QY 841 aatgggtcttcctcgaaggaggttcctcctgacagaagaagcgaattctcttaccat 900
DB 841 AATGGTCTTCTTTCGAGGAGATGCTTCCCTGTCAGAACGATTTCTGTTTACCTCCAT 900
QY 901 ggttgcaagacatccttgaggttgaacctgaacgggcaaatcagaagccttctctcat 960
DB 901 GGTTCAGACGATCTTGAATTTTGAACCTGACGGGCAATATCAGAGCTTCTTCCATAG 960
QY 961 caatcgctcaatgaagtgagagcaagtaactgctcctcgccgaagcccgacttaag 1020
DB 961 CAATCGCTCAATGAGATGAGACCAAGTTCACTGCTCTGCGCAAGCCGCACTCAG 1020

QY 1021 gaccgaagcccatatggtgcttgcggcgagactagcaacaaccacaacacagagatgg 1080
DB 1021 GACCAAGGCCCATATGTTGGCTTGGGCGACGTAGCAACACCAACCAACGACGATGG 1080
QY 1081 ctgagatcgaatttggggagaaaaaataacaggaattagaaccagaagcttaca 1140
DB 1081 CTGAGATCGAATTTGGGGAGAAAAAGAAATAACAGGAATTAAGAACACAGATCTACA 1140
QY 1141 cagtcgaactccaacttattgtaagaagtttggtagaacttcaaaacaaataatttc 1200
DB 1141 CAGTCGAACCTCAACTTTATGTTAAGAGTTTGTGATGAACCTCAAAAACAAATAATCT 1200
QY 1201 aagtgaagacactataaaggaaattgataataaagaaagggtgttccaggtaactc 1260
DB 1201 AAGTGAAGACCTATAAAGAAATTTGTGATTAATAAATAAAGGTTTCAGGGTAACCTC 1260
QY 1261 aacttcgggaccagtgcaaaacaaattcaaccctccatcgtggcagatattgycgg 1320
DB 1261 AACTTTCGGGACCCAGTCCAAACAAATTTCAATCCCTCCATCTGCGCAGATATGTGCG 1320
QY 1321 gttgtcccccagagacatgacacagagatagccttgaagtgtagatcattgttcag 1380
DB 1321 GTTGTCCCCAGACATGGCACAGAGATAGCTTGAAGGTGAGATGATGTTGGCAG 1380
QY 1381 attacaagaagtaatgattcatttgggtggtgcgaagaacaaagtcgaagcagtgctta 1440
DB 1381 ATTACCAAGTAATGATTCAATGTTGTGCGCAAGAACAAAGTCAACACACAGTGTTC 1440
QY 1441 actgaagaagaagatgagaacataacaaagcccatccctcctggaagaagaatccaga 1500
DB 1441 ACTAAGAAAGAGATGAGAACATCAACAGGCCATCCCTCGGAAGAACATCCACAGGA 1500
QY 1501 ataaacattacaagtgagctatccattcattggtgctcctgtgtcctggtgcttgc 1560
DB 1501 ATAAACATTACAAGGTGGCTATTCATTGGTGGCTGTGTGCTGTGCTGTGCTGCTG 1560
QY 1561 atgggagctcttgcagccttgaagaagaagaagaagaagtcgtagtgcagcg 1620
DB 1561 ATGGGAGCTTGTGACGCCCTTGAAGAAAGAAAGAAAGAAAGAAAGTCCGTATGCTCAG 1620
QY 1621 gaggcctcagaagaagcagctggttgaagcagataataatcccttgcagaacatgca 1680
DB 1621 GAGGCTCAGAAAGAAAGACTGTTGGAAGCAGATTAATTCCTTTCAGACATCATCA 1680
QY 1681 gctgaattcacatcagctatgataatgagaagagatgacacaagaattagatcctac 1740
DB 1681 GCTGAATTACCATCATGATGATTAATGGAAGCAGATGACCAAAAGTTGATCTCATC 1740
QY 1741 acaagtgataatgagagtttaa 1761
DB 1741 ACAAGTGAATATGGCAGCTTAA 1761

RESULT 2
AX118820 1620 bp DNA linear PAT 11-MAY-2001
LOCUS AX118820 Sequence 5 from Patent WO0129219.
DEFINITION AX118820
ACCESSION AX118820
VERSION AX118820.1 GI:14035776
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1620)
AUTHORS Donoho, G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and Sands, A.T.
TITLE Human cub-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 5 26-Apr-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..1620

Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 240 ggggagagagctgggtgtagtgctgtagacacagctagtcacatgacagtagtgacacat 299
    |||||||
Db 1 GGGGAGAGAGCTGGGTGATGGCTGTGACACCTAGTGAATATCAGGATAGTGGACAAAT 60
    |||||||

QY 300 gacatcagaatltaccgggaacctaccacacatcttccggaagaagacacatc 359
    |||||||
Db 61 GACATCTAGAAATTTCCCGGAGCTACCCCAATACACTGTGTCGAAAAGACAAATAC 120
    |||||||

QY 360 agtaccaaagggaagaagactgattctgaggttggagatttgatcgaatccagac 419
    |||||||
Db 121 AGTACCAAGGGGAAACACTGATCTGAGGTTGGAGATTTGGATTCGAATCCCAAC 180
    |||||||

QY 420 ctgtctctgactatctctctcacaagctctcagaatcgaatcgaatcgaatcgaat 479
    |||||||
Db 181 CTGTCTCTGACTATCTCTCTCTACACAGCTTTCAGATCAATATGATCCATATCTGG 240
    |||||||

QY 480 aagtaagactgttcccaagaagactctgttgaacaagaagtaaacgctccgcttga 539
    |||||||
Db 241 AAGTATGACTGTTCACCAAGAACTCTGTGACACAGTGAAGTAAACCGTCCGTTTGA 300
    |||||||

QY 540 gagtgatccacacatctctgcccggggttcttctgactatgacgagcagacacatcc 599
    |||||||
Db 301 GAGTGGATCCCATTTCTGGCCGGGTTTCTGCTGACCTATGCGACAGGACCAATCC 360
    |||||||

QY 600 agatttaataaagtttggaaagagctagaccatatttgaaagaagaatacagaacatt 659
    |||||||
Db 361 AGATTTAATTAACATTTTGGAAACAGCTAGCCATTTATTTGAAGACAGAAATACAGCAATT 420
    |||||||

QY 660 ctgcccacgtggttgaagaacgtagcagagacacatcttctggaatcgtgtagatga 719
    |||||||
Db 421 CTGCCACGCTGGTGTAGAGAGCTAGACAGACATTTCTGGGAATATGTGTAGATGATA 480
    |||||||

QY 720 tagagataaccttattatgtgcaaaagctgcacatcgaatcgaagaataatctgtatgaact 779
    |||||||
Db 481 TAGAATACCTCTTTATTTATGTGCAAAAGCTGCACATCCATCAGAAATATTTGCGATGAATC 540
    |||||||

QY 780 aggtggccagatcagtgctgtctcagcgaagaaggaatcagtgatcgaagaagattctgcgc 839
    |||||||
Db 541 AGGTGGCCAGATCAGTGTGCTTCAAGCGCAAAAGGATCAAGTCAATATGAAAGGATTTCTGCG 600
    |||||||

QY 840 caatgggttcttctcgaaggaatggttccctgtcagacaagaagcattctgttaccctca 899
    |||||||
Db 601 CAATGGTGTCTTTTCGAGGAGATGTCTCCCTGCACAGCAAGCATTTCTGTATTCCTCAA 660
    |||||||

QY 900 tgggtgcagcagatctctgtgaggttgaacacctgacggggaacatcgaagttcttccctatg 959
    |||||||
Db 661 TGGTGCAGCAGATCTTTGAGTTTGAACCTGACGCGCAAAATCAGAGCTTCTTCTCTCATG 720
    |||||||

QY 960 gcaatcgttcaatgagatgtagaacaagctcagctgtcctcgtgccaagcccgactca 1019
    |||||||
Db 721 GCAATCGGTCATGAGAGTGAAGACCAAGTTCACTGCTCTGCGCAAGCCGACTTCA 780
    |||||||

QY 1020 ggaaccaagcccaatcagtggtctggcgacagtagcaacaacaacaagaagagtg 1079
    |||||||
Db 781 GGACCAAGGCCCATCATGGGCTTCGGGGAGCAGTAGCAACCAACCAACGAGAGTG 840
    |||||||

QY 1080 gctggagatcgaattggggggaagaagaataaacaagaatcgaagaacagagatc 1139
    |||||||
Db 841 GCTGGAGATCGATTTGGGGGGAGAAAAGAAAATAACAGAGATTAAGACACAGATCTAC 900
    |||||||

QY 1140 acagtcgaactcaacttattatgttaagaagtttgatgaactcacaataaacaataatc 1199
    |||||||
Db 901 ACAGTCGAACCTCAACTTTATGTTAAGAGTTTGTGATGAACCTTCAAAAACAAATATTC 960
    |||||||

QY 1200 taagtggaaagacctataaagaagattgtgaataaagaagaagtggttcagggtaactc 1259
    |||||||
Db 961 TAAAGTGAAGACCTATAAAGAAATGTGAATATGAAGAAAAGGCTTTCAGGGAATCTC 1020
    |||||||

QY 1260 taactttggagccagtgcaaaacatcaccctccatcgttggccagatgataatgctg 1319
    |||||||
Db 1021 TAACCTTTCGGGACCAAGTGCAGAAACAAATTTTCCTCCATCGTGGCCAGATATGTGCG 1080
    |||||||

```

```

QY 1320 gtttcccccagacatgacacacagagagatagccttgaagtgtagctcatgtgttcca 1379
    |||||||
Db 1081 GGTGTCTCCCCAGACATAGGACGACGACAGATAGCTTAAAGTGTGAGCTCATTTGGTGCA 1140
    |||||||

QY 1380 gattacaagaatgaatcattcattgtgtgtgctgcaagaagaagtaaacacagtgcttc 1439
    |||||||
Db 1141 GATTACCAAGAGTAATGATTTCATGTGTGTGGCGCAAGACAAAGTCAACGACGATGTTTC 1200
    |||||||

QY 1440 aactaagaagaagaatgaagaatcacaagaagccatccctcgtggaagaagaatccacag 1499
    |||||||
Db 1201 AACTAAGAAAGAAATAGACAAATACACAGGCCATCCCTCGAAGAAACATCCACAGG 1260
    |||||||

QY 1500 aataacattacaacgctgtgctatccatctgtgtcctctgttgcctgtgtgttgcctg 1559
    |||||||
Db 1261 AATAAACATTATCAACGCTGCTATTCATGTGTGTCTCTCTGTTGTCTGCTGTGTGCTG 1320
    |||||||

QY 1560 aatgggaatcttgcagcctttagaagaagaagaagaagaagtcgataagatcagc 1619
    |||||||
Db 1321 AATGGGATCTTTTCAGCTTTAGAAAGAAAGAAAGAAAGAAAGTCCGTATGATCAGC 1380
    |||||||

QY 1620 ggaagctcagaanaacagactgttgaagaagatgaatcattcccttgcaggaatcagtc 1679
    |||||||
Db 1381 GGAGGCTCAGAAACAGACTGTTGGAAGCAGATTAAATATCTCTTGCACAGACATCAGTC 1440
    |||||||

QY 1680 agctgaattlacatcagctatgataaagaagaagatgacacaagaagtagatctcat 1739
    |||||||
Db 1441 AGCTGAGTTTACATCACTAGCTATGATATGAGAAAGATGACACAAAAGTTAGATCTCAT 1500
    |||||||

QY 1740 cacaatgatatgagcaggttaa 1761
    |||||||
Db 1501 CACAAGTGAATATGACAGGTTAA 1522
    |||||||

RESULT 4
AX118816 1464 bp DNA linear PAT 11-MAY-2001
LOCUS AX118816
DEFINITION Sequence 1 from Patent WO0129219.
ACCESSION AX118816
VERSION AX118816.1 GI:14035774
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1464)
AUTHORS Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
Sands,A.T.
TITLE Human cnd-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 1 26-Apr-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..1464
location/Qualifiers
BASE COUNT 440 a 303 c 359 g 362 t
ORIGIN
Query Match 83.1%; Score 1464; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 1.3e-269;
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


Db 121 ACTGCTCTTGTGACTATCTTCTTCCACCAGCTCTTCAGATCAATATGCTCATCTGT 180
OY 478 ggaagtagactgttcccaagaactcttgttgaaccaagtgaagtaaccgttc 537
Db 181 GGAAGTGTGACTGTCTCCCAAGAACTCTTGTGAACAAGTGAAGTAAACCGCTCTT 240
OY 538 ggaagtagatcccaacttctgscggggttcttctgtagccta tggagcaagacat 597
Db 241 GAGAGTGTGATCCCACTTCTGCGCGGGGTTTTTGTGTGACATATGCGACAGACAT 300
OY 598 ccagatttaataacatgttggaaagagtagccattatttgaagacagaatacaat 657
Db 301 CCAATTTAATATACATCTTTGGAACGAGTAGCCATTTATTGAAGACAGAAATACAGCAA 360
OY 658 ttctgcccaagctggtttaaagacgttagcggagacattcttggaaata tggtagatga 717
Db 361 TTCTGCCAGCTGTGTGTAAAGAGTAGAGAGACATTTCTGGAAATATGTAAGATGCA 420
OY 718 taaagataactcttatttctgcaaaagctgccatgcaaggaataattgtatga 777
Db 421 TATAGATATACCTTTTATTGTGCAAAAGTGCATCCATCAGGAATATGCTGATGAA 480
OY 778 ctgagtgccagatcagtggtctcagcgcaaaaggaatcagatatagaaggattctg 837
Db 481 CTAGGTGCGAGATCACTGTGCTTCAAGCGAAAGGATGATATGAAGGATTTCTG 540
OY 838 gcaaatgtgtcttctcgaagagatgtgtccctgtcacaagaagcaattctgttacttc 897
Db 541 GCCAATGTTTCTTCTGAGAGATGTTCCCTGTGACACAAGGATTTCTGTTTACCTCC 600
OY 898 aatgtgtcagcagatccttgaatttgaacctgaacgggaatcagagcttctctca 957
Db 601 AATGTTGACAGCATCTTGAAGTTTGAACCTGACGCGCAAAATCAAGACTTCTCTCA 660
OY 958 tggcagctcagtagaaggaatggagaccagttcactgtctctgccaagccgactt 1017
Db 661 TGGCAGCTCGGTCAATGAGATGAGACCAAGTTTCACTGCTCTGCGCCCAAGCCGACTT 720
OY 1018 caggacaaagccatcaltatgggtctcgcgagcagtagacaacaacaacagcagag 1077
Db 721 CAGGACCAAGGCCATCATGTGGCTTGGGGGAGTAGACAAACCAACCAACGAGAG 780
OY 1078 tggctggagatcgatttggggggaagaaataaacaaggaattagaccacagatct 1137
Db 781 TGGCTGAGATCATTTGGGGGAGAAAGAAATTAACAGGATTAAGACACAGATCT 840
OY 1138 aaccagtcgagactcaacttltatgaagcttggtagaactcaaaacaataat 1197
Db 841 ACACAGTCGACTTCAACTTTATGTTAAGATTTGTGATGAACCTTCAAAAACAATTAAT 900
OY 1198 tctaagtgaagacctataaaggaattgttaataatgaagaaaggltgttcagggtaac 1257
Db 901 TCTAAGTGAAGACCTATTAAGGAATTTGCAATATGAAGAAAGAGTGTTCAGGGTAA 960
OY 1258 tctaacttccgggaacccagtgcaaaacaatttcaatccctcccaatcg tggccagatattg 1317
Db 961 TCTAATCTTTCGAGCCAGCCAGTCAAAACAAATTTATCCCTCCATCGGCGACATATG 1020
OY 1318 cgggtgttcccccagacatgagccagagagatgacttgaagtgtagcttcatgttgc 1377
Db 1021 CGGGTGTCTCCCGACATATG6GACCGAGAGATAGCTTGAAGGTGAGCTCATTTGGTTTC 1080
OY 1378 cagattacaagaagtatcatctatgtgtgtgcgcaagaagaagtcagaagaccagtgt 1437
Db 1081 CACATTAACAAAGTATATGTTGTTGCGCGCAACAAAGTCAAAAGCCACAGTGT 1140
OY 1438 tcaactaagaagaagatgagacaatcacaaagccatccctcggaaagaanaacatccaca 1497
Db 1141 TCAACTAAGAAAGATAGACAAATCAACAAGGCCATCCCTCGAAGAAACATCATACA 1200
OY 1498 ggaataaactaacaggttgatctatctatctatctgtgtctgtgtgtgtgtgtgt 1557
Db 1201 GGAATTAACATTAACAAGGTGCTATTCATTTGGTGTCTCTTGTGTCTCTGTTGTGCT 1260

OY 1558 ggaatgggagctcttccagcctttaaagaagaagaagaagagcttcgtatgatca 1617
Db 1261 GGAATGGGATCTTTTGCAGCTTTAGAAAGAAAGAAAGAAAGAGTCCGTATGATCA 1320
OY 1618 gggagagctcagaagaacagactgttggagcagattaaatcccttccagaaatcag 1677
Db 1321 GCGAGGCTCAGAAACAGACTCTTGGAAAGCAATTAATCTCTTCCAGACATCG 1380
OY 1678 tcaagctgaattacacatcagctatgataatgaagaagagatgacacaaagttagctc 1737
Db 1381 TCAGCTGAGTTTACATCAGCTATGATATGAAAGAGATGACACAAAAGTTAGATCTC 1440
OY 1738 atcaagaatgatatggcaggttaa 1761
Db 1441 ATCAAGATGATATGCGAGGTAA 1464

RESULT 5
AC099419/c
LOCUS
DEFINITION
AC099419
VERSION
KEYWORDS
SOURCE
ORGANISM

AC099419 249429 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus chromosome Mcsi clone CH230-8N22, WORKING DRAFT
SEQUENCE, 42 unordered pieces.
AC099419 GI:17942157
HTG: HTGS_PHASE1; HTGS_DRAFT.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 249429)

REFERENCE
AUTHORS

Muzny D.M., Adams C., Adio-Oduola B., Ali-oshman F.R., Allen C.,
Alshrooks S.L., Amaratunga H.C., Are J.R., Banks T., Barbara J.,
Benton J., Blimage K., Blankenburg K., Bonnin D., Bouck J.,
Bowie S., Brileva M., Brown M., Brown M., Bryant N.P., Buhay C.,
Burck P., Burkett C., Burrell K.L., Byrd N.C., Carroon T.F.,
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Davy Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
Deem A.L., Ding Y., Dinh H.H., Douthwaite R.J., Draper H.,
Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C.,
Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
Garza N., Gill R., Gorrell J.H., Guevara M., Gunaratne P., Hale S.,
Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,
Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C.,
Hollins B., Homs F., Howard S., Huber J., Hulyk S., Hume J.,
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,
Louiseged H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,
Ma J., Maheshwari M., Mapa P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhney E., Mcleod M.P., Mesdor M.,
Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nokinokko S.,
Oguth M., Okunou G., Oragunye N., Oyedro R., Pace A., Payton B.,
Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,
Quiles S., Ren Y., Rives M., Rojes A., Rojibokan I., Roife M.,
Ruiz S., Savery G., Scherer S., Scott G., Shen H., Shoshart N.,
Sisson I., Sodergren E., Sonakke T., Sparks A., Stanley H.,
Stone H., Sutton A., Swalek A., Tabor P., Tamerisa A., Tamerisa K.,
Tang H., Tansey J., Taylor C., Taylor T., Telford B., Thomas N.,
Thomas S., Usmani K., Vasquez L., Vera V., Villalón D., Vinson R.,
Wall R., Wang S., Ward-Moore S., Warren R., Washington C.,
Watlington S., Williams G., Williamson A., Wleczky R., Wooden S.,
Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,
Weinstock G. and Gdbbs R.

TITLE
JOURNAL
Direct Submission
unpublished

Db 4370 GTGCAACGATCAGTAT-GACCCATATGTGGAGTTGGCTGTTCACAGAACCTCTGGC 4312

Qy 509 tgaacaaaggaagtaacccgtccgttgagatgagatcccaatttcgcccggggtt.568

Db 4311 TGAATCTCAAGAGCGACTGTGGCTTCAGAGGATGATCAATTCCTCGCCGGGCT 4252

Qy 569 ttctgtcactatgcagcagcagccatcagatattacaatggtttggaagagta 628

Db 4251 TTTCTGTCAGCCATTCATGCAATGATCATCCAGATTAACTGTTTGAATGAGCA 4192

Qy 629 gccattattgaagacagacaataatctcgcacagctggtgtgtagaagacgtagag 688

Db 4191 GTCAATTATTCACAGAAAAATACAGCCAGTCTGCCCCAGCTGTGTGAGACATAGCAA 4132

Qy 689 gagcattctcgggaatatgtagatgatatagagactcttatttgcaaacgtg 748

Db 4131 GGGCAATTCCTGGGAAATCAAAAGATGCTACTAGATAGCGCTTATGTGCAAACTG 4072

Qy 749 ccatcatgcagaaataatgctgtagaactaggtgcccagatca-----gtgtgct 800

Db 4071 ACATTCATGACAGGATCATCTCAATGAACTAGGTGGCTACGACACCTGCTCAGAGCT 4012

Qy 801 tcagcgaagaagagtcagtgatataagagatctgcgcaatggttcttctgagga 860

Db 4011 TCAAGTCAAGAGGTTAAGTCACTCTGAGGGAATCGTGGCCAAATGCTGTGCTCCAA 3952

Qy 861 tgcctcctgcagacaagcagattcctgttactccaatg 902

Db 3951 TGGTTCCTGTGACA-AAGTCATTTCTGTTTACATCCCAAG 3911

RESULT 6
HSDJ92C8 131060 bp DNA linear PRI 21-AUG-2000
LOCUS Human DNA sequence from clone RPI-92C8 on chromosome 6, complete
DEFINITION
ACCESSION AL132671
VERSION AL132671.20 GI:9685189
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 131060)
AUTHORS Tyromans,A.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:7228316.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RPI-92C8 is from the library RPI-1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.bufileo.edu/>
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RPI-92C8
It may be shorter because we sequence overlapping sections only
once, except for a 100 base overlap.
The true left end of clone RPI-179P9 is at 13095 in this sequence.
The true right end of clone RPI-94G16 is at 100 in this sequence.
location/Qualifiers
1. 131060

FEATURES
source
repeat_region
/note="Aluub repeat: matches 11. .293 of consensus"
425. .513
/note="L2 repeat: matches 2562. .2648 of consensus"
514. .856
/note="MER1B repeat: matches 1. .337 of consensus"
857. .1162
/note="L2 repeat: matches 2246. .2562 of consensus"
1977. .2048
/note="L2 repeat: matches 2629. .2704 of consensus"
3004. .3273
/note="Aluub repeat: matches 18. .287 of consensus"
4127. .5300
/note="CpG island"
/evidence="not_experimental"
4614. .4755
/note="71 copies 2 mer CC 59% conserved"
4621. .4707
/note="3 copies 29 mer 81% conserved"
6477. .6564
/note="2 copies 44 mer 95% conserved"
7144. .7280
/note="FAM repeat: matches 22. .163 of consensus"
7828. .7888
/note="MSTa repeat: matches 1. .61 of consensus"
8038. .8120
/note="L2 repeat: matches 2465. .2551 of consensus"
8223. .8531
/note="Alusd repeat: matches 1. .305 of consensus"
8639. .9030
/note="MER31a repeat: matches 1. .417 of consensus"
9455. .9539
/note="MER57-internal repeat: matches 1252. .1347 of
consensus"
9543. .9743
/note="MER57-internal repeat: matches 2005. .2220 of
consensus"
9732. .9907
/note="MER57-internal repeat: matches 2456. .2629 of
consensus"
9900. .10720
/note="MER57-internal repeat: matches 2777. .3637 of
consensus"
10709. .11327
/note="MER41-internal repeat: matches 3972. .4590 of
consensus"
11328. .11618
/note="Alusg repeat: matches 1. .292 of consensus"
11619. .12914
/note="MER41-internal repeat: matches 4590. .4554 of
consensus"
12915. .13144
/note="L1Pa5 repeat: matches 5914. .6143 of consensus"
13145. .13805
/note="MER41-internal repeat: matches 4553. .5213 of
consensus"
13824. .14458
/note="MER41B repeat: matches 1. .633 of consensus"
15901. .16301
repeat_region

```

repeat_region /note="AluY repeat: matches 1. .311 of consensus"
18848. .19305
/note="L1PB1 repeat: matches 5685. .6152 of consensus"
repeat_region /note="MER21A repeat: matches 684. .933 of consensus"
19336. .19575
/note="MER21A repeat: matches 684. .933 of consensus"
19570. .19759
/note="MER21B repeat: matches 326. .531 of consensus"
20528. .20618
/note="MIR repeat: matches 48. .146 of consensus"
20926. .21059
/note="L1MC1 repeat: matches 6189. .6327 of consensus"
21601. .21760
/note="AluY repeat: matches 144. .304 of consensus"
22605. .22747
/note="MER91A repeat: matches 28. .159 of consensus"
22956. .23435
/note="L1MD repeat: matches -1. .475 of consensus"
23463. .23624
/note="L1MD repeat: matches 860. .1022 of consensus"
23626. .23692
/note="MER7A repeat: matches 283. .346 of consensus"
23693. .24058
/note="THE1C repeat: matches 1. .369 of consensus"
24059. .24264
/note="MER7A repeat: matches 63. .283 of consensus"
24547. .24602
/note="28 copies 2 mer aa 71% conserved"
24622. .25110
/note="L1M4 repeat: matches 2209. .2702 of consensus"
25129. .26287
/note="L1M4 repeat: matches 3174. .4259 of consensus"
26288. .26589
/note="L1M4 repeat: matches 1. .302 of consensus"
26590. .26692
/note="L1M4 repeat: matches 3067. .3174 of consensus"
26694. .26816
/note="L1M4 repeat: matches 2707. .2830 of consensus"
26862. .28116
/note="L1MD1 repeat: matches 4968. .6213 of consensus"
28118. .28233
/note="4 copies 29 mer 73% conserved"
28958. .29218
/note="L1ME3A repeat: matches 5319. .5594 of consensus"
29294. .29442
/note="L1ME3A repeat: matches 5744. .5894 of consensus"
29793. .29901
/note="U6 repeat: matches 1. .107 of consensus"
32842. .32990
/note="MIR repeat: matches 74. .230 of consensus"
33015. .33293
/note="AluY repeat: matches 3. .305 of consensus"
33294. .33430
/note="AluY repeat: matches 1. .138 of consensus"
33881. .33949
/note="MIR repeat: matches 67. .142 of consensus"
34337. .34618
/note="AluY repeat: matches 2. .283 of consensus"
34753. .35058
/note="AluY repeat: matches 2. .296 of consensus"
37633. .37939
/note="AluY repeat: matches 3. .309 of consensus"
38218. .38311
/note="MER81 repeat: matches 1. .104 of consensus"
38453. .38920
/note="L1M87 repeat: matches 5694. .6171 of consensus"
39205. .39278
/note="MIR repeat: matches 117. .193 of consensus"
39919. .40220
/note="AluY repeat: matches 3. .302 of consensus"
41119. .41144
/note="13 copies 2 mer ac 100% conserved"
41234. .41534
/note="AluY repeat: matches 1. .298 of consensus"

```

```

repeat_region 41920. 42076
/note="L2 repeat: matches 2187. .2344 of consensus"
repeat_region 42077. 42367
/note="AluY repeat: matches 12. .297 of consensus"
repeat_region 42368. 42399
/note="L2 repeat: matches 2156. .2187 of consensus"
repeat_region 44012. 44103
/note="MER33 repeat: matches 1. .92 of consensus"
repeat_region 44551. 44785
/note="MIR repeat: matches 2. .249 of consensus"
repeat_region 44928. 45276
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 46018. 46250
/note="L2 repeat: matches 2498. .2742 of consensus"

Query Match 14.6%; Score 257; DB 9; Length 131060;
Best Local Similarity 100.0%; Pred. No. 5,5e-39;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggattcgtgtgacggagcagactgacgacgctcccgagcgccgctcgtccgagag 60
|||||
DB 4994 ATGGATTTCGTGTGACGGAGCAGACTGACGACCCCTCCCGCGCGCTCGTCCGACAG 4935
|||||
QY 61 gagcgagcccgcccgccgagcagctcgagctcgagatccgctcgaaggagcgagcttgc 120
|||||
DB 4934 GAGCGAGCCCGCGCGCGAGCTGCGCTCGGATCCGTGAGGAGGAGCGGAGCTTGCC 4875
|||||
QY 121 aagctggagcccgagggggtatgltgcccggagcccgagcgagcgacatcgacg 180
|||||
DB 4874 AAGCTGGAGCCCGAGGGGATGTTGCCCCGCGCGCGCGCGCGCGCGCGCGCGCG 4815
|||||
QY 181 gctgcggagcgagcgctcctcgtcttgctgctcgagtcgagtcgagcccgctcgagtcag 240
|||||
DB 4814 GCTGCGGCGCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4755
|||||
QY 241 gcggagagagctggtga 257
|||||
DB 4754 GCGGAGAGAGCTGGTGA 4738
|||||

RESULT 7
AF387548 2642 bp mRNA linear ROD 11-NOV-2001
LOCUS Mus musculus strain ICR endothelial and smooth muscle cell-derived
DEFINITION neurophilin-like protein (Esdn) mRNA, complete cds.
ACCESSION AF387548
VERSION AF387548.1 GI:16902436
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2642)
REFERENCE Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.
ESDN, a Novel Neurophilin-like Membrane Protein Cloned from Vascular
Cells with the Longest Secretory Signal Sequence among Eukaryotes,
Is Up-regulated after Vascular Injury
J. Biol. Chem. 276 (36), 34105-34114 (2001)
JOURNAL 11447234
PUBMED 2 (bases 1 to 2642)
REFERENCE Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Sasayama,S., Matsumori,A., Honjo,T. and Tashiro,K.
Direct Submission
TITLE Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
JOURNAL University Graduate School of Medicine, Yoshida Konoe-cho,
Sakyo-ku, Kyoto 606-8501, Japan
FEATURES
source Location/Qualifiers
1. .2642
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
1. .2642
gene

```



```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1.2339
/contig="Esdn"
12.2321
/contig="Esdn"
/codon_start=-1
/product="endothelial and smooth muscle cell-derived
neuropilin-like protein"
/protein_id="AAL30180.1"
/db_xref="GI:16902439"
/translation="MASRAPLRAARSPDGGRAAPATGAPLPSAGMCLPGRHNS
SSRPRLILLILLPDAQAGKGGCGTIVPESGTSINYPHTVPSNVCWEIRV
KGRERIRKRGDIDSDYCHGNKLTGKIGFISGRTSGKGLIOMNOSIESGS
ETIVLPMGSHAGRGRLASVYIDKODLITCDITVSNFPEPESKCPRGCLLPRAE
ISGTPHGYDSDSPLCMAGIHAGVSDVLGQISVISKPTYESLANLISMGV
LSTSLFPTSGCYTGMEGV IADPQITASSVLEWIDHMGQNSKPEKARLRPG
PPMAFATDEHOWLIDLNEKKITGIVTSTLIEHNYVSAVLYSDQGRWYV
REGAADKIFQGNKDYHKDVRNNELPILARLRVNPVQMOOKIAMKYLSCFTL
KGRLPKTOPPPRNSNKLNTYHPKLGARPKETOLGPRSRNDLPLDPOITAPD
VKNTTTPSTTKVLAALVYPPVVALTLLILVCAHMRNRKKAESTYDLPNHD
RAGMKGVKQLPAKSYHEHEVTVKSNSEVSHUSPREVTIVLQADSAEYAPLVGI
VGTJHORSTKPEBGEKASVADLPYNAPOEYHAYAEPLPTGPRYATPIVMDNG
HSTASVGLPSTSTFRAGNOPALVGYNTLLSRDSCSSGQAOYDTPKGGKPAAPAE
ELVYOVOSTOEAAGAGRDEKDFAFKETL"
BASE COUNT      612 a      606 c      594 g      527 t
ORIGIN
Query Match      12.5%; Score 220.4; DB 10; Length 2339;
Best Local Similarity 50.1%; Pred. No. 3.6e-32;
Matches 688; Conservative 0; Mismatches 641; Indels 43; Gaps 4;

```

```

OY 628 agccattattggaacagacatacagaacattctgccagcttggtgtaagacgtatga 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 600 TCTAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGCTGCTGCTTTT 659
OY 688 ggaacattcttggaatagttagatgatatagatacctcttatttgacaagct 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 660 GCTGAATATCTGGAACGATTCCTCATGATATAGAGATTTCTTACCGCTGTATGAGCT 719
OY 748 gcatccatctcagaataattgcttgatgaactagtagtgccacagcttgctcagcg 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 720 GAATCCATCCAGAGTAGTGTGATGATGCTGGTGGTCCAAATACGCTTGATTAAC 779
OY 808 aagagacacgcgatgataagaaggatcttgccaatggtgtcttcgaaggatctcc 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 780 AAGAGCACCACATATTAAGAAAGTCTTTGGCCAAATGTCACTTCCATGAGGGATAC 839
OY 868 ctgcaagaacagatctcgtttacccatagttgtcagaagatcccttgagtttga 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 840 TTATCTACGAGCTGTTTACATTTAAGACAAAGTGTCTATGGACTTAAAGGATGAG 899
OY 928 cctgacgg-----caatcagaagctcttcctcatgagcagctgcacatgag 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 900 TCAGGTGTGATCCGCCATVCCCGATACAGCATCTGTACTGGAGTGGACTGACAC 959
OY 976 agtgaagacaaatctcactgctcctgycacaagccagactcagaagccagatca 1035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 960 ATGGGGCAGAGGAACACTGTGAAACCCGACAAAGCCAGCTGAGAAACCGGGGCTCC 1019
OY 1036 tggagcttcggcgagcagtagcaacaacacacagcagaagtggtgtgagatgattg 1095
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1020 TGGGCTCTCTTTGGCACTGATGAGCAATC-----AGTGGCTGCAAAATTTGACCTT 1067
OY 1096 ggggagagaaaagaataacaggaattagagcacaagatctacacagtcgaactcaac 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1068 AATTAAGAGAGAAAGATTAACAGGATGTGAACCACTGTGATCCCTGATAGAGCAAAAT 1127
OY 1156 tttaattgaagatttgatgtaacttcaaaaataattcctaagtggaaacctat 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1128 TACTATGTGCTGCTGCTACAGATTTCTGTACAGTACGATGGCAGAAATGAGTGTGAC 1187
OY 1216 aaaggaatttgataataatgaagaagaagtggttcagagtaaccttaactctggagccca 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1188 AGAGAGCTGGTGGCTGAGACAAATATTTCAAGGAACAAATTTATCAACAGAT 1247
OY 1276 gtgcaacaacattcatccctccatcgttgccagatattgcygggtgtgtccccaagaca 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1248 GTTGTATTAATCTTTTGGCCAAATTAATGCAAGTTTATAGAGTGAACCTGTCCAG 1307
OY 1336 tggcacaagagatagccttggaaggtgagatcatgtgttgccaagataaac 1387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1308 TGGCACAGCAAAATTTGCCATGCAAGTGAATTTGCTGGAGTGTGATGACTGC 1359

```

RESULT 11

```

HS94G16      100808 bp      DNA      linear      PRI 23-NOV-1999
LOCUS      HS94G16
DEFINITION      Human DNA sequence from PAC 94G16 on chromosome 6q21. Contains.
ACCESSION      285999
VERSION      285999.1 GI:2326514
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

REFERENCE 1 (bases 1 to 100808)

AUTHORS

TITLE

JOURNAL

Submitted (08-AUG-1997) Chromosome 6 Project Group
 (http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquer@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk
 On Aug 14, 1997 this sequence version replaced gi:1842200.

IMPORTANT: This sequence is the entire insert of clone 94G16. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre chromosome 6 mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6/>. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 94G16 is at 1 in this sequence. The true right end of clone 94G16 is at 100808. 94G16 is from the library RPCII constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.

For further details see <http://Dacpac.med.buffalo.edu/>. Location/Qualifiers

1. 100808

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="6q21"

/clone="RPI-94G16"

/clone_lib="RPCI-1"

2. 109

/note="AluSg repeat: matches 193. .300 of consensus; incomplete repeat"

repeat_region 386. .684

repeat_region 828. .1128

repeat_region 1511. .1913

repeat_region 2069. .2366

repeat_region 2560. .2589

repeat_region 4018. .4220

repeat_region 4380. .4673

repeat_region 5738. .5790

repeat_region 6052. .6137

repeat_region 7850. .8066

repeat_region 9495. .9676

repeat_region 11587. .11890

repeat_region 12013. .12188

repeat_region 12310. .12500

repeat_region 13051. .13336

repeat_region 13800. .14096

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region 16740. .16877

repeat_region

repeat_region 17538. .17591

repeat_region 18151. .18287

repeat_region 18554. .18722

repeat_region 20152. .20294

repeat_region 21524. .21613

repeat_region 21616. .21915

repeat_region 21916. .22025

repeat_region 21983. .21992

repeat_region 22793. .22848

repeat_region 23337. .23527

repeat_region 26788. .27074

repeat_region 27487. .27787

repeat_region 28466. .28524

repeat_region 28475. .28537

repeat_region 28538. .28843

repeat_region 29096. .29432

repeat_region 29433. .30977

repeat_region 30980. .31322

repeat_region 32594. .32810

repeat_region 34663. .34860

repeat_region 34906. .35188

repeat_region 36667. .36969

repeat_region 37045. .37345

repeat_region 39246. .39818

repeat_region 41305. .41604

repeat_region 42255. .42299

repeat_region 42702. .43037

repeat_region 44291. .44516

repeat_region 47198. .47261

repeat_region 49349. .49557

repeat_region 49728. .50016

repeat_region 50103. .50104

repeat_region 50117. .50111

repeat_region 50140. .50235

repeat_region 51950. .52234

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557

repeat_region 52255. .52557


```

misc_feature 97164..142616
/Note="assembly_name:Contig10"
BASE COUNT 43950 a 26378 c 28311 g 43375 t 602 others
ORIGIN

Query Match 12.4%; Score 218; DB 2; Length 142616;
Best Local Similarity 100.0%; Pred. No. 1.5e-31;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 ggtgcagcagatcccttgagtttgacactgacgggaacacagagcttctctcatg 960
|||||
Db 141261 GGTTCAGCAGATCTCTGAGTTTGAACTGACGGCAAAATCAGAGCTTCTCTCATG 141202
|||||

QY 961 cagtcggtcaatgagatgagacacagttcagctcctgccaagcccgcttcag 1020
|||||
Db 141201 CAGTCGCTCATGAGAGTGGAGACCAAGTTCTGCTGCTGCGCAAGCCCGCTTCAG 141142
|||||

QY 1021 gaccaagggcccatcatgagcttcgagcagatgacacacacacacacagagatg 1080
|||||
Db 141141 GACCAAGGCCCATCATGAGGCTTCGGGCGACAGTAGCAACACCAACACAGAGACTGG 141082
|||||

QY 1081 ctggagatcagtttgaggaggagaaaaaataacag 1118
|||||
Db 141081 CTGGAGATCGATTGTTGGGAGAAAAAGAAATTAACAG 141044
|||||

RESULT 13
AC099428/c 171300 bp DNA linear HTG 20-DEC-2001
LOCUS
DEFINITION
Rattus norvegicus clone CH230-101K6, *** SEQUENCING IN PROGRESS
AC099428
VERSION
AC099428.2 GI:17973372
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 171300)
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barberia,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvad,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,U., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lousised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newtonson,J., Newtonson,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokewo,S.,
Ogun,M., Okunolu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,D., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolle,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Slison,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Yinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 171300)
Worley,K.C.
Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16930835.
-----
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GHUB
Center clone name: CH230-101K6
-----
Summary Statistics
Assembly program: Phrap, version 0.990329first call to
findPhrapList
Consensus quality: 138590 bases at least Q40
Consensus quality: 14764 bases at least Q30
Consensus quality: 15580 bases at least Q20
Estimated insert size: 152896; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 11285: contig of 11285 bp in length
* 11286 11385: gap of unknown length
* 11386 116925: contig of 5540 bp in length
* 116926 17025: gap of unknown length
* 17026 20478: contig of 3453 bp in length
* 20479 20578: gap of unknown length
* 20579 26556: contig of 6078 bp in length
* 26557 26756: gap of unknown length
* 26757 33936: contig of 7180 bp in length
* 33937 34036: gap of unknown length
* 34037 38968: contig of 4932 bp in length
* 38969 39068: gap of unknown length
* 39069 43965: contig of 4997 bp in length
* 43966 44065: gap of unknown length
* 44066 48815: contig of 4750 bp in length
* 48816 48916 48915: gap of unknown length
* 48917 51807: contig of 2892 bp in length
* 51808 51907: gap of unknown length
* 51908 56443: contig of 4536 bp in length
* 56444 56543: gap of unknown length
* 56544 59759: contig of 3216 bp in length
* 59760 59859: gap of unknown length
* 59860 64034: contig of 4175 bp in length
* 64035 64134: gap of unknown length
* 64135 68670: contig of 4536 bp in length
* 68671 68770: gap of unknown length
* 68771 75473: contig of 6703 bp in length
* 75474 75573: gap of unknown length
* 75574 79858: contig of 4285 bp in length

```

```

* 79859 79958: gap of unknown length:
* 79959 83831: contig of 3873 bp in length
* 83931: gap of unknown length
* 83932 86749: contig of 2818 bp in length
* 86750 86849: gap of unknown length
* 86850 89988: contig of 3139 bp in length
* 89989 90088: gap of unknown length
* 90089 93114: contig of 3026 bp in length
* 93115 93125: gap of unknown length
* 93125 95428: contig of 2214 bp in length
* 95429 95528: gap of unknown length
* 95529 99646: contig of 4118 bp in length
* 99647 99746: gap of unknown length
* 99747 102546: contig of 2800 bp in length
* 102547 102646: gap of unknown length
* 102647 104912: contig of 2266 bp in length
* 104913 105012: gap of unknown length
* 105013 108083: contig of 3071 bp in length
* 108084 108183: gap of unknown length
* 108184 111047: contig of 2864 bp in length
* 111048 111147: gap of unknown length
* 111148 112820: contig of 1673 bp in length
* 112821 112920: gap of unknown length
* 112921 115880: contig of 2660 bp in length
* 115881 115680: gap of unknown length
* 115681 117534: contig of 1854 bp in length
* 117535 117634: gap of unknown length
* 117635 119548: contig of 1914 bp in length
* 119549 119648: gap of unknown length
* 119649 121877: contig of 2229 bp in length
* 121878 121977: gap of unknown length
* 121978 124058: contig of 2081 bp in length
* 124059 124158: gap of unknown length
* 124159 126116: contig of 1958 bp in length
* 126117 126216: gap of unknown length
* 126217 128515: contig of 2299 bp in length
* 128516 128615: gap of unknown length
* 128616 130603: contig of 1988 bp in length
* 130604 130703: gap of unknown length
* 130704 132213: contig of 1510 bp in length
* 132214 132313: gap of unknown length
* 132314 134064: contig of 1751 bp in length
* 134065 134164: gap of unknown length
* 134165 136099: contig of 1935 bp in length
* 136100 136199: gap of unknown length
* 136200 137441: contig of 1242 bp in length
* 137442 137541: gap of unknown length
* 137542 139039: contig of 1498 bp in length
* 139040 139139: gap of unknown length
* 139140 140911: contig of 1772 bp in length
* 140912 141011: gap of unknown length
* 141012 142584: contig of 1573 bp in length
* 142585 142684: gap of unknown length
* 142685 144313: contig of 1629 bp in length
* 144314 144413: gap of unknown length
* 144414 146560: contig of 2147 bp in length
* 146561 146660: gap of unknown length
* 146661 148283: contig of 1623 bp in length
* 148284 148383: gap of unknown length
* 148384 150257: contig of 1874 bp in length
* 150258 150357: gap of unknown length
* 150358 151846: contig of 1489 bp in length
* 151847 151946: gap of unknown length
* 151947 153174: contig of 1228 bp in length
* 153175 153274: gap of unknown length
* 153275 155485: contig of 2211 bp in length
* 155486 155585: gap of unknown length
* 155586 156894: contig of 1309 bp in length
* 156895 156994: gap of unknown length
* 156995 158140: contig of 1146 bp in length
* 158141 158240: gap of unknown length
* 158241 159705: contig of 1465 bp in length
* 159706 159805: gap of unknown length

```

Query Match 9.0%; Score 159; DB 2; Length 171300;
 Best Local Similarity 79.1%; Pred. No. 2.9e-20;
 Matches 189; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

QY 253 ggtatgctgctgagaccctcagtgactatcagatgctgcaaatcagatcgaat 312
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9559 GGTATGCTGCTGAGCGGCACATGTCATCTCAGACGTCGACAAATGACATCTTAAGAAAT 9500
QY 313 tatcccgagactaccccaatcacactgtctgcgaaagaacattacagactcaagaaggg 372
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9499 TATCCAGGACTTACCCCAATTAACCGTGTCGAAAGATCATACAGTCCGAAAGGGG 9440
QY 373 aaagactgattctgaggttggagatttgatatacgaaatccagacctgtgtctgac 432
      || ||||| ||||| ||||| ||||| || || || || || ||||| ||||| |||||
DB 9439 AAGAGACTTATTCAGAGCTGGCGGCAATTCAGACATTCAGACAGACCTCGCTCCGAC 9380
QY 433 tatctcttcacacactcttcagatcaataatgctacatcctgtaagatgactgt 491
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9379 TATCTCCTTTACGACGACGACATGATGTAAGCAAGCGGATATGGCATCTGT 9321

```

RESULT 14

AC026646 142616 bp DNA linear HTG 17-NOV-2000
 LOCUS
 DEFINITION Homo sapiens chromosome 6 clone RP11-117013, WORKING DRAFT

SEQUENCE 7 unordered pieces.

AC026646.5 GI:10998991
 HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS
 human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eulhethia; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 142616)

Waterson, R. H.

Unpublished

2 (bases 1 to 142616)

Waterson, R. H.

Direct Submission

Submitted (22-MAR-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Oct 25, 2000 this sequence version replaced gi:9958143.

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_MH0117013
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; Version 0.99019
 Consensus quality: 139293 bases at least Q40
 Consensus quality: 140128 bases at least Q30
 Consensus quality: 140609 bases at least Q20
 Insert size: 14100; agarose-fp
 Insert size: 142145; sum-of-contrigs

D	301	TCAAGATGGAATTCACATTCTTTCGGACGGGAAATTTTGCGCTCACAATCCGTATTACATA	360
OY	596	atc--cagatttaatacatglttbgaaacagctlagca-ttatttgaagacaagaatacacg	653
D	361	AACAAAGACTAATTAATCTGTGTTGGACACTGCATGCCAATTTTTTTGGAAACCTGGATTACG	420
OY	654	caaatctc--ccagactggtttgaaagacgtagcaasgagcaatcttcggataatgtag	712
D	421	TAACTACTGCCCCACGGCTGGTGTGCTCTTCCTTTTTCGTAAGATATCTGGAAACAATTCCTC	480
OY	713	atggatalaagaatatacctctttaattgtgcgaaaactgcacatcalbgcagaataatgtcgt	772
D	481	ATGGATATAGAGATTCCTCCGCAATTTGTGCATGGCTGGTGGATGCAAGATACTGTCAA	540
OY	773	atgaactagtggtgccagatacgaatgctgtccacgcgaanaagggatcagtataaaggaa	832
D	541	ACACGTTGGGGGGCCAATATAGTGTGTAATTAGTAAGAAGTATTCCTATTATGAAGTT	600
OY	833	tctcgaccatacgtgtctcttcgcgaagatggtlccctgtccagacaagaatctct--	889
D	601	CTTTGGCTAACACAGTCAACTCTGTGGTGGACACTTGTACTACAGCTTTTTCAT	660
OY	890	ttaacctccaatgyltgcagcagacatccttgagtttgaacctgaacgggacaatcagagct	949
D	661	TTAAGACAAAGTGATGTTTATGGAACAACCTGGGNAATGACACTGTGGTATCGCGCATCT	720
OY	950	cttccctaigtgcgtgtaataatgagatgagagccaagtltcaatgctccctgcgaag	1009
D	721	CAAATACACGATCACTACTGTGCTGTGAGTGGACTGACACTGACACACAGGCAAGAGAAAGTTGGA	780
OY	1010	cccgaactcaggaaccaagcccataatggtcgttcggcgacagtagcaacaaccaacaac	1069
D	781	AACCCAAAAGAACCCAGGCTGAAAAAACCTGGACCCTGTGGGTGCTTTGGCACATGATG	840
OY	1070	cacgaagatgctcgtgaatcogatcttgggggagaanaaataacagaaatciagacca	1129
D	841	AATACCACTGTTCAATATGATTGTGANTAGGAAAGAAATACAGGCAATTATMACCA	900
OY	1130	caggaactacacacagtcgaacttcaactlltaigtlaagagtttgtlgaacticaaaa	1189
D	901	CTGGATTCACACATGTTGAGC-ACAAATTACTATGTGCTGCCTACAGAAATCCTGTACAGTG	959
OY	1190	acaaataatctcaagtggagaagaactataaag-gaatgtgtaataatgaaagaagaagtgctt	1248
D	960	AGGATGGGCAGAAATATGCACTGTGTACGAGAAAGGCTGTGTCGAGCAACATATAGTTATTT	1019
OY	1249	cagggtaactctaactcttcggagc---cgaagtcaaaaaaatllcatacctcccatcg	1305
D	1020	CAGGAAACAAAAGCAATTATTCACAGAGATGTTGCTGTAATTAACTTTGGCACCAATTAAT	1079
OY	1306	gccagatatgtgcgggtgtgtccccagacatlygcaccaagaagagatcettlgaagtgag	1365
D	1080	G-CACGTTTATTAGAGTGAATCCTTACCAATGGCAGCAGAAATTTGCCATGAANAATGGAG	1138
OY	1366	cctcaatgtgttcgagatta 1384	
D	1139	CTGCTGGATGTCAGTTTA 1157	

Search completed: July 9, 2002, 01:13:21
Job time: 6038 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 8, 2002, 23:26:13 ; Search time 1913.71 Seconds
(without alignments)
12419.927 Million cell updates/sec

Title: US-09-691-344A-3

Perfect score: 1761

Sequence: 1 atgggattcgctgcgggca.....caagtgatcgacagttaa 1761

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	540.8	30.7	1934	11	AK016485	AK016485 Mus muscu
2	513.8	29.2	2835	11	AK014521	AK014521 Mus muscu
3	470.6	26.7	872	10	BG245910	BG245910 602358725
4	444.8	25.3	883	10	BG975603	BG975603 602845306
5	434.4	20.2	466	9	AM390343	AM390343 RC4-ST018
6	355.4	19.0	434	9	AM483656	AM483656 55586 MAR
7	335.4	15.5	406	10	BF740455	BF740455 RC2-HB000
8	273.2	15.5	975	10	BF144325	BF144325 601787144
9	252.2	14.3	257	9	BE068835	BE068835 MRO-BT037
10	249.2	14.2	260	10	BF331266	BF331266 MRO-BT037
11	242.2	13.8	474	9	AA050442	AA050442 m112e10.r
12	228.4	13.0	464	9	BB860022	BB860022 BB860022
13	212.2	12.0	451	10	BF086926	BF086926 RC3-HT023
14	209.4	11.9	406	9	AA042037	AA042037 m104n12.r
15	203.2	11.5	896	10	BF139411	BF139411 601785285
16	198	11.2	412	9	AI935615	AI935615 w098f03.x
17	189.8	10.8	318	10	H32466	H32466 EST107591 R

18	185.2	10.5	482	10	BF871665	BF871665 OVO-ET015
19	179	10.2	219	9	BE068836	BE068836 MRO-BT037
20	160.8	9.1	914	10	BI413107	BI413107 602990241
21	155.6	8.8	295	9	BS585422	BS585422 BS585422
22	150.6	8.6	643	9	BB179946	BB179946 BB179946
23	148.8	8.4	894	10	BG257635	BG257635 602377506
24	146.2	8.3	608	9	BB657747	BB657747 BB657747
25	138.8	7.9	670	12	B57216	B57216 CIT-HSP-200
26	136.4	7.7	667	9	AI586285	AI586285 AI586285
27	134	7.6	664	9	AM383889	AM383889 OY3-HT036
28	129.8	7.4	639	9	AI048059	AI048059 DKF2P586F
29	128.2	7.3	587	9	AM383890	AM383890 QV3-HT036
30	127.8	7.3	459	10	N46066	N46066 YY35905.r1
31	127.2	7.2	466	9	AI750960	AI750960 cm06908.x
32	125.2	7.1	693	10	BF733384	BF733384 nael1a12.
33	121	6.9	986	10	BI080802	BI080802 602878574
34	120	6.8	463	10	BF777080	BF777080 288892 MA
35	119.2	6.8	692	10	BF732861	BF732861 nael16d03.
36	114	6.5	908	10	BG178371	BG178371 602330141
37	113.2	6.4	299	10	BF350722	BF350722 OY3-HT036
38	108.8	6.2	270	9	BI177683	BI177683 BI177683
39	102.2	5.8	703	9	AM070902	AM070902 xa31b09.x
40	102	5.8	324	10	H84766	H84766 ys70e05.r1
41	101	5.7	408	10	BF825585	BF825585 CM1-HN001
42	97.8	5.6	578	9	AA122907	AA122907 mq23a10.r
43	96	5.5	745	9	AL577036	AL577036 AL577036
44	95.8	5.4	506	9	AI048089	AI048089 DKF2P586H
45	94.4	5.4	927	9	AL552613	AL552613 AL552613

ALIGNMENTS

RESULT 1	AK016485	1924 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK016485				
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4931429A04:putative CUB domain containing protein, full insert sequence.				
ACCESSION	AK016485.1	GI:12855243			
VERSION	AK016485				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:4931429A04.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2 (sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata, K., Itoh, M., Alizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuzaki, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				

[illegible][illegible]

Query Match	20.2%	Score 355.4	DB 9	Length 434
Best Local Similarity	90.3%	Pred. No. 8.9e-66		
Matches 380	0	Mismatches 41	Indels 0	Gaps 0

```

FEATURES      High quality sequence: 406.
              Location/Qualifiers
source        1..406

```

```

/db.xref="taxon:9606"
/clone.lib="HB0006"
/seq.stage="Adult"
/seq.stage="Adult"
/site_2: Smal; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      135 a      80 c      101 g      90 t
ORIGIN
Query Match      19.0% Score 335.4; DB:10; Length 406;
Best Local Similarity 94.9%; Pred. No. 1.7e-61;
Matches 369; Conservative 0; Mismatches 16; Indels 4; Gaps 2;
QY 1372 ggttcgcagattacacaaggaatgatcatctgtgtgtgcgcgaagacaagtcacaagcacc 1431
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GGAGGCATAGTTATTCATGAGGTATGATTCATGATGAGTGTCG -GCAAGACAAAGTCAAAAGCAC 59
QY 1432 agtgccttaactaagaagaagaatgagcgaatcacaaaggcccatccctctggaagaaca 1491
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 AGTGCTCAACTCAAAAGAAAGATGACAAATCAACAAAGCCCATCCCTCCGAAACAAACA 119
QY 1492 tccacaaggaataaacaattacacaacgctgacatcttcattgtgtcctctgtctcgtgtg 1551
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 TCCACAGGAATAAATCATTTACAAGGCTGCTATTCATGTGTGCTCTCTTGATGCTGTG 179
QY 1552 ttctgctgaatgagggaatctttgcagcctttgaaagaagaagaagaagatccgat 1611
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 TTTGCTGGAATGCGGATCTTTGCACCCCTTTAGA --AAGAAGAAAGAAAGAAAGTCCGCTAT 236
QY 1612 ggaatcagaggagggcccaaaaacagacagctgttggaaagacagatataatcccttgccaga 1671
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 GGATGAGGAGAGAGCGCTCCAAACAAACAGACTGTGGAAAGCAGATTAATATTCCTTGCCACGA 296
QY 1672 catcagtcagctgagtttaccatcacgtacgtatgataatgaaagagatgacacaaaagtta 1731
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 CATCAGTCAGCTGAGTTTACCATCTGCATGATATGATGAGGAGATGACACAAAAGTTTA 356
QY 1732 gatctcatcacaaagtatgataatgccaagtta 1760
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 GATCTCATCACAAAGATGATGTGACAGATTA 385
RESULT 8
BF144325 975 bp mRNA linear EST 24-OCT-2000
LOCUS 601787144F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4014778 5',
DEFINITION mRNA sequence.
ACCESSION BF144325
VERSION BF144325.1 GI:10983365
XREF EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 975)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: L1AM9259 row: h column: 11

```

High quality sequence stop: 642.
Location/Qualifiers
1. .975
/organism="Mus musculus"
/strain="C57BL 6J"
/db_xref="taxon:10090"
/clone="IMAGE:4014778"
/clone_1id="NCI-CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; transgenic model MMT-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT
248 a . 240 c 299 g 188 t

ORIGIN

Query Match 15.5%; Score 273.2; DB 10; Length 975;
Best Local Similarity 82.4%; Pred. No. 4, 2e-48;
Matches 361; Conservative 0; Mismatches 73; Indels 4; Gaps 4;

Oy 466 ggtccatactctgtagaagatgactgtgtcccaagaactctgttgaacacaagtgaagta 525
Db 200 GGTCCATATTAGTGGAGATTGGGCTGTCCCAAAGAACTCCGGCTGAACTCAAGAAAGT 259

Oy 526 accgtccgctttgtagaagtgagatcccaatttcgtcgcgagggtttttgttaactatg 585
Db 260 ACTGTCTCTTCAMAGAGGATCTACATTTCTGGCCGGGCTTTCTGCTACCTTACGCC 319

Oy 586 agcagcgacacatcccgatattataacagtttctgtagcagcagtagaaccattattgaagaca 645
Db 320 AGCACTGACCAATCCAGATTTATATACCTGTTTGGAAAGCAGCAGCAGCATTTATTTGAGGAA 379

Oy 646 gaatacagaacaattctgcgcagctggtgtgtagaagcgttaagaagaacaattctgggaat 705
Db 380 AATATACACCAATTTCTGCCACGCTGGCTGTAGAGACATAGACGAGGAAATTTCTGGGAA 439

Oy 706 atgtagatgtagatagagatacactcttattgtgtcaagaagctgcacatcatgacaggaata 765
Db 440 ACAAAGAGATGG-TTACAGAGATATACCTTTATTGTGCCAAGCTG-CATCCAGCGAGGATG 497

Oy 766 attgctgtatgaactatggtggcca-gataagtgtgcttaacgcgcaaaaggtatcagtcgata 824
Db 498 ATCACAGATGAACGTAGGGGCGCACACCATCAACTTGGCTTCAGAGCACAGGAAATGACTACTA 557

Oy 825 tgaagaggaattctgcgcgaattgttctcttcgtgagagatggttccctgttaacagcagat 884
Db 558 TGAAGAGACTCCTGG-CAATGGCGGTCTCTCCCGGCAATGTTCTTTGTGCGAAAACCATTT 616

Oy 885 tctgttaactctccaatg 902
Db 617 TCTTTTTCACAAACCCGAG 634

RESULT 9
BE068835 257 bp mRNA linear EST 09-JUN-2000
LOCUS BE068835
DEFINITION MR0-BR0374-220300-001-F07 BR0374 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE068835
VERSION BE068835.1 GI:8413485
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 257)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Bionesi, M.R.,
Nagal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,
GOLDMAN, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

TITLE Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0-BT0374-220
300-001-f07&t3=2000-03-22&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 257.
Location/Qualifiers
1. 257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0374"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 87 a 48 c 60 g 62 t
ORIGIN
Query Match 14.3%; Score 252.2; DB 9; Length 257;
Best Local Similarity 98.8%; Pred. No. 9.5e-44;
Matches 254; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1090 gatttgggggagagaaataacaggaattagagacacagagctacacagtcgaac 1149
|||||
DB 1 GATTGGGGGAGAGAAATAACAGCAATTAGCACACAGCTACGCGATGCGAAG 60
QY 1150 ttcacattatgttaagaggttttgatgaactcaaaaacataatctaaagtggaag 1209
|||||
DB 61 TTCACATTTTATGTAAAGGTTTGTGATGAACCTCAAAAACAATTAATCTAAGTGAAG 120
QY 1210 acctataaggaattgtgataatgaagaaagtgcttcaaggtaacttaacttcg 1269
|||||
DB 121 ACCTATAAGGAATTGTGAATATGAAGAAAGGTGTTCAAGTAACCTTAACCTTTCGG 180
QY 1270 gaccacagtcgaaacaatttcacccatcgctgagcagatagtcgaggttccccc 1329
|||||
DB 181 GACCCGTCACAAACGATTCATCCCTCCCATCGCGCAGATATCGCGGTGTGCCCC 240
QY 1330 cagacatggcaccagag 1346
|||||
DB 241 CAGACATGGCACCAGAG 257
RESULT 10
BF31266 260 bp mRNA linear EST 22-NOV-2000
LOCUS BF31266
DEFINITION MR0-BT0374-220300-001-h01 BT0374 Homo sapiens CDNA, mRNA sequence.
ACCESSION BF31266
VERSION BF31266.1 GI:11302014
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 260)

AUTHORS Dias Neto,E., Garcia Correa,R., Vejnovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,U.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0-BT0374-
220300-001-h01&t3=2000-03-22&t4=1)
Seq primer: puc 18 forward.
Location/Qualifiers
1. 260
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0374"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 90 a 49 c 61 g 60 t
ORIGIN
Query Match 14.2%; Score 249.2; DB 10; Length 260;
Best Local Similarity 98.8%; Pred. No. 4.2e-43;
Matches 251; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1093 ttgggggagagaaataacaggaattagagcacaagagctacacagtcgaacttc 1152
|||||
DB 7 TTGGGGGAGAGAAATAACAGCAATTAGCACACAGCATCTACACAGTCAACTTC 66
QY 1153 aacttctatgaagagtttgtatgaacttcaaaaacataatctaaagtggaagacc 1212
|||||
DB 67 AACTTTATGTAAAGGTTGTGATGAACCTCAAAAACAATTAATCTAAGTGAAGACC 126
QY 1213 tataaaggaatttgaataatgaagaaagtgcttcaaggtaacttaacttcggagc 1272
|||||
DB 127 TATAAAGGAATTGTGAATATGAAGAAAGGTGCTCAGGTAACCTTAACCTTTCGGGCC 186
QY 1273 ccaagtgcacaaacatttaccctccatcgctggccagagatagtgcgggttccccag 1332
|||||
DB 187 CCAAGTGCACAAACATTTATCTCCCTCCCATCGTGCCAGATATGTGCGGTTGTCCCCAG 246
QY 1333 acatggcacacagag 1346
|||||
DB 247 ACATGGCACACAGAG 260
RESULT 11
AA050442 474 bp mRNA linear EST 09-SEP-1996
LOCUS AA050442
DEFINITION m112e10.r1 Soares mouse embryo NM0213.5 14.5 Mus musculus CDNA
clone IMAGE:475914 5', mRNA sequence.
ACCESSION AA050442
VERSION AA050442.1 GI:1530113
KEYWORDS EST.

	SOURCE	house mouse.
	ORGANISM	Mus musculus
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 474)
	AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubucque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelsing,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
	TITLE	The MASHU-HMI Mouse EST Project
	JOURNAL	Unpublished (1996)
	COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:286658 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 468. Location/Qualifiers 1..474 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:475914" /clone_lib="Soares mouse embryo NDME13.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: pRTT3-Pac (Pharmacia) with a modified polylinker; site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer 15' TGTTACCAATCTGAAAGTGGGAGCGGCCGAATTATTTTTTTTTTTTTTTTTT T 3], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Patima Bonaldo."
	BASE COUNT	95 a 120 c 161 g 98 t
	ORIGIN	
Query Match	13.8% Score 242.2; DB 9; Length 474;	
Best Local Similarity	81.0%; Pred. No. 1.5e-41;	
Matches 294; Conservative 0; Mismatches 68; Indels 1; Gaps 1;		
OY	159	cggcgccgagcactgtgcggcgctgccggcgcccttcgtcttgcctcgcgcgt 218
Db	113	CGGCGGACCACTGGGTACCAGCGCGCGGTGGGCACAGTCTTGGGCGTCGTTCCGCGGT 172
OY	219	ctccgccccgcctcgcgtcgcagcgaggagacgtgatgtcgttggaacacctagtac 278
Db	173	GTHGCTCCGCTCCGGTTGACGAGGGAGACTGGTGATAGGTGTGGCACATAAGTAC 232
OY	279	ttaaccagatgtgtgcacaatgatcatctaagaattatoccggaactaacccaatacac 338
Db	233	CTCTCAGACAGGTGGCAACAATGACATCTTAAGAATTATCCAGGACTTAACCCAATTACAC 292
OY	339	tgtttgcgaagaacaatttacagfcaaccaaaagggaaaagactgtattcgaagtgggaga 398
Db	293	TGCTGTGAAAAGTATCATCAAGTCCCACAAAGGGAAGAGACTTATTTCTGAAGTTGGGAGA 352
OY	399	tttgtatcatcgaatcccagacctgtcttcgtactatcttccttaaccaaccttttaga 458

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	353	TTTGAACATTTGAGTCCAAAGACCTGGCCTTCTGACTATCTTCCTTCAGACATGCAAGCA	412					
Qy	459	ttaatatggtccatctactgtggaagcatgctcttcccaagaagctctgttgaacacaag	518					
Db	413	TGAGTATGGTCCATATTTGGAGATTGGGCTG-TCCCAAGAACATCCGGCTGAACTCAAA	471					
Qy	519	tga 521						
Db	472	CGA 474						
RESULT	12							
LOCUS	BB860022							
DEFINITION	BB860022	RIKEN full-length cDNA enriched, pooled cell lines Mus musculus						
ACCESSION	BB860022							
VERSION	BB860022.1	GI:17101476						
KEYWORDS	EST.							
SOURCE	house mouse.							
ORGANISM	Mus musculus							
REFERENCE	BB860022	464 bp mRNA linear EST 26-NOV-2001						
AUTHORS	BB860022	RIKEN full-length cDNA enriched, pooled cell lines Mus musculus						
JOURNAL	unpublished (2001)							
COMMENT	Contact: Yoshihide Hayashizaki							
	Laboratory for Genome Exploration Research Group, RIKEN Genomic							
	Sciences Center(GSC), Yokohama Institute							
	The Institute of Physical and Chemical Research (RIKEN)							
	1-7-22 Shuho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan							
	Tel: 81-45-503-9222							
	Fax: 81-45-503-9216							
	Email: genome-res@gsr.riken.go.jp,							
	URL: http://genome.gsc.riken.go.jp/							
	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh							
	, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.							
	Normalization and subtraction of cap-trapper-selected cDNAs to							
	prepare full-length cDNA libraries for rapid discovery of new							
	genes. Genome Res. 10 (10), 1617-1630 (2000)							
	wagii, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E.,							
	Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura							
	, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and							
	Hayashizaki, Y.							
	RIKEN integrated sequence analysis (RISA) system-384-format							
	sequencing pipeline with 384 multicapillary sequencer. Genome Res.							
	10 (11), 1757-1771 (2000)							
	Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara							
	, Y., and Hayashizaki, Y.							
	Computer-based methods for the mouse full-length cDNA							
	encyclopedia: real-time sequence clustering for construction of a							
	nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)							
	Please visit our web site (http://genome.gsc.riken.go.jp) for							
	further details.							
	e mouse tissues.							
FEATURES	Location/Qualifiers							
SOURCE	1..464							
	/organism="Mus musculus"							
	/db_xref="taxon:10090"							
	/clone_id="GA30011K05"							
	/note="pooled cell lines ; (cell_line=CRJ-1751 WEHI 16d) ,							
	(cell_line=RCB-0464 Meth-4) , (cell_line=RCB-0545 OHTA) ,							

```

BASE COUNT      95 a      120 c      153 g      96 t
ORIGIN
(
  (cell_line=RCB-0559_K-1_F1), (cell_line=RCB-1283_B16
  melanoma), (cell_type=B cells, cell_line=CRU-1702_MH1_231
  ), (cell_type=Leydig cells, cell_line=CRU-2065_MTC-1),
  (cell_type=Nullipotent stem cell, cell_line=CRU-2070_NE),
  (tissue_type=Bladder, cell_line=RCB-0544_MBT-2),
  (tissue_type=Bone marrow, cell_type=Stroma cell,
  cell_line=CRU-2028_SR-4987), (tissue_type=Colon,
  cell_line=RCB-0549_Cle-H3), (tissue_type=Kidney,
  cell_line=CCU-142_RM6), (tissue_type=Submandibular gland,
  cell_line=CRU-1734_SCA-9_Clon15), (strain=BAB/C,
  cell_type=B cells, cell_line=CRU-1669_BC11_Clon13_20-3B3
  ), (strain=C3H, tissue_type=brain, cell_line=CRU-1443
  BC3H1)"

```


FEATURES

1000

2011

WYOTING

query

Qy 21

!

3304

500

SECRET

REVIEWS

TAYLOR

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 8, 2002, 23:34:13 ; Search time 78.07 Seconds
(without alignments)
5540.676 Million cell updates/sec

Title: US-09-691-344A-3

Perfect score: 1761

Sequence: 1 atgggattcggtggtggggca.....caagtgatgcaaggttaa 1761

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92.8	5.3	6909	2	US-08-804-196-1 Sequence 1, Appli
2	92.8	5.3	6909	2	US-08-658-340-1 Sequence 1, Appli
3	92.8	5.3	6909	3	US-08-746-111-26 Sequence 26, Appli
4	81.6	4.6	7032	4	US-09-324-867-1 Sequence 1, Appli
5	80.6	4.6	6585	3	US-08-746-111-4 Sequence 4, Appli
6	78.4	4.5	4629	2	US-08-484-891-7 Sequence 7, Appli
7	78.4	4.5	4670	3	US-08-717-294-41 Sequence 41, Appli
8	78.4	4.5	4999	4	US-09-470-618-14 Sequence 14, Appli
9	78.4	4.5	4999	4	US-09-364-862-14 Sequence 14, Appli
10	78.4	4.5	5035	2	US-08-882-083-1 Sequence 1, Appli
11	78.4	4.5	5035	3	US-08-558-107-1 Sequence 1, Appli
12	78.4	4.5	5035	3	US-09-243-539-1 Sequence 1, Appli
13	78.4	4.5	6999	1	US-08-276-594A-1 Sequence 1, Appli
14	78.4	4.5	7056	1	US-08-121-202-1 Sequence 1, Appli
15	78.4	4.5	8241	6	5171844-1 Patent No. 5171844
16	78.4	4.5	8967	1	US-08-366-851A-1 Sequence 1, Appli
17	78.4	4.5	9009	1	US-07-864-004B-3 Sequence 3, Appli
18	78.4	4.5	9009	1	US-08-251-937A-3 Sequence 3, Appli
19	78.4	4.5	9009	1	US-08-212-133A-1 Sequence 1, Appli
20	78.4	4.5	9009	1	US-08-474-503-1 Sequence 1, Appli
21	78.4	4.5	9009	2	US-08-670-707A-1 Sequence 1, Appli
22	78.4	4.5	9009	4	US-09-037-601-1 Sequence 1, Appli
23	78.4	4.5	9009	5	PCT-US93-03275-3 Sequence 1, Appli
24	78.4	4.5	9354	1	US-08-683-839B-2 Sequence 2, Appli
25	78.4	4.5	11933	1	US-09-470-618-13 Sequence 13, Appli
26	78.4	4.5	11933	4	US-09-364-862-13 Sequence 13, Appli
27	78.4	4.5	11933	4	US-09-364-862-13 Sequence 13, Appli

28	58	3.3	2303	2	US-08-480-229C-9 Sequence 9, Appli
29	58	3.3	2303	2	US-08-659-235C-9 Sequence 9, Appli
30	57.6	3.3	1384	1	US-07-607-538C-1 Sequence 1, Appli
31	57.6	3.3	1384	2	US-08-162-402B-1 Sequence 1, Appli
32	57.6	3.3	1934	3	US-08-162-402B-7 Sequence 7, Appli
33	54.4	3.1	4451	3	US-08-717-294-42 Sequence 4, Appli
34	53.8	3.1	318	2	US-08-480-229C-19 Sequence 19, Appli
35	53.8	3.1	318	2	US-08-659-235C-19 Sequence 19, Appli
36	53.4	3.0	7493	1	US-08-212-133A-7 Sequence 7, Appli
37	53.4	3.0	7493	1	US-08-474-503-5 Sequence 5, Appli
38	53.4	3.0	7493	4	US-08-670-707A-5 Sequence 5, Appli
39	53.4	3.0	7493	4	US-09-037-601-5 Sequence 5, Appli
40	53.4	3.0	7493	5	PCT-US94-13200-5 Sequence 5, Appli
41	52.8	3.0	4334	2	US-08-670-707A-38 Sequence 38, Appli
42	52.8	3.0	4334	4	US-09-037-601-38 Sequence 38, Appli
43	52.8	3.0	6402	4	US-08-670-707A-36 Sequence 36, Appli
44	52.8	3.0	6402	4	US-09-037-601-36 Sequence 36, Appli
45	52.2	3.0	1780	2	US-08-480-229C-11 Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-08-804-196-1
Sequence 1, Application US/08804196
Patent No. 5874256
GENERAL INFORMATION:
APPLICANT: Bertina, Rogier
TITLE OF INVENTION: A method for diagnosing an increased risk for thrombosis or a genetic defect causing thrombosis and kit for use with the same.
TITLE OF INVENTION: risk for thrombosis or a genetic defect causing thrombosis and kit for use with the same.
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESS: Akzo No. 5874256el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,196
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/454,353
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gornley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 948-9751
TELEFAX: (301) 948-9751
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-804-196-1

Query Match 5.3% Score 92.8; DB 2; Length 6909;
Best local Similarity 51.0% Pred. No. 1.2e-13;
Matches 248; Conservative 0; Mismatches 232; Indels 6; Gaps 1;
QY 898 aatgttcgaagatccttgatttgaacctgacggaatcgaagcttcttcctca 957

Db	6280	AATGGATGTTCCACACCCCTGGGATATGGAAATGGAAAGATGGAAAACACCAATTCACA	6339
OY	958	tgagcagtcgylcaalvgagvglygagacaaagtlcaactgylctcctgycgaagccggaactt	1017
Db	6340	GCTTCCTTCGTTTAAAGAAATCTTGTTGGGAGATTACTGGAGAACCTTCCTGGTCCGCTCG	6399
OY	1018	caggagcaagagccatcatatgagtgcttcggcgagcagtaagaaacacacaacacagagag	1077
Db	6400	AATCCCAAGGAGCGTGTGAATTCGCTGGCAAGCCAAAGCAACCAACAAATTAAGC----	6453
OY	1078	tgagctgtagatcgatitgvgvggagaaagaaataacagagatcaggacacagagatct	1137
Db	6454	TGGCTAGAAATTAATGATCTACTCAGATGACAAAGATACCGGCAATTTATTAACACAGGGCTGC	6513
OY	1138	acacagtcgaactlcaacttitaatgyltaagagtlttgylgtagacttcaaaacaacatcat	1197
Db	6514	AAGCTCTGTCCTCTGTAAGATGATGTAAGAGCGTATACATCCATACAGTAGAGCAGGA	6573
OY	1198	tctaaagvgagagacctataaagaaatlgtaataatgaagaaagvgtlcttcagvgtaac	1257
Db	6574	GTTGGAATGGAACCATTAACAGCGTGAATCCTCCATGCTGTGACAAACATTTTGAAGGAAT	6633
OY	1258	tctaacttcggagaccagcagtgcaaaacaatltcatccctcccatcgvgtgcagatatlgty	1317
Db	6634	ACTTAATACCAAGGACATGTGGAAGAACTTTTTCACACCCCCCAATATATTTTCCAGGTTTATC	6693
OY	1318	cggagltgccccagaagatgycacacagagatagccttgaagvgtgaagctcatlgtgtgc	1377
Db	6694	CGTGCTATTCCTTAACATGAGATCAAGATTAATACACTTCGCTGGAACTCTTTGGCGCT	6753
OY	1378	cagatc 1383	
Db	6754	GATATT 6759	

RESULT 2
 US-08-658-340-1
 Sequence 1, Application US/08658340
 Patent No. 5910576
 Patent No. 5910576 5861489
 GENERAL INFORMATION:
 APPLICANT: Bertina, Rogier
 APPLICANT: Roltsma, Pieter
 TITLE OF INVENTION: A method for diagnosing an increased
 TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
 TITLE OF INVENTION: thrombosis and kit for use with the same.
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Akzo No. 5910576 5861489el Patent Department
 STREET: 1300 Piccard Drive, Suite 206
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/658,340
 FILING DATE:
 PRIOR APPLICATION DATA: '
 APPLICATION NUMBER: US 08/454,353
 FILING DATE: 06-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Gormley, Mary E.
 REGISTRATION NUMBER: 34,409
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 948-7400
 TELEFAX: (301) 948-9751
 INFORMATION FOR SEQ ID NO: 1:

```

:      SEQUENCE CHARACTERISTICS:
:          LENGTH: 6909 base pairs
:          TYPE: nucleic acid
:          STRANDEDNESS: both
:          TOPOLOGY: unknown
:      MOLECULE TYPE: CDNA
US-08-658-340-1

```

Query Match	5.3%	Score 92.8	DB 2	Length 6909
Best Local Similarity	51.0%	Pred. No. 1.2e-13		
Matches 248	Conservative	0	Mismatches 232	Indels 6
			Gaps	1
QY 898	aatggttcgaagagatcctctgagtttggaaacttgacgggccaatcagagcttcttccca	957		
Db 6280	AATGGATGTTCCACACACCCCTGGGTATGGAAATGGAACATATGMAAACAGCAAAATCACA	6339		
QY 956	tgagcagtcggtcaatgatgagatggagaccagttcacttgctctctgagccagcccgact	1017		
Db 6340	GCTTTCTTCCTTTAAGAAATCTTGTTGGGGAGATTACTGGGAACCCCTTCGTCGCCGTCTG	6399		
QY 1018	cagagcccaaggcccatcaatgagcttcggcgacgtagaagaacaacacacagagag	1077		
Db 6400	AATGCCCAAGGAGCGTGTGATGCTCTGGCAGGCCAAGCCAAACACATATAGC-----AG	6453		
QY 1078	tgagcggagatccgatttgggggagaaaaaagaataaacaagagatcagagccacagatct	1137		
Db 6454	TGGCTTAGAATTTGATCTACTCAAGATCAAGAAAGATACCGCATTTATATACACAGGCTGC	6513		
QY 1138	acacagtcgaacttcaacttctatgtctaaagatttggatgaacttcaaaaaacaataat	1197		
Db 6514	AAGTCTCTGTCTCTGAAATGTATGTAAAGACACTATACCATCCACTACAGTGAGCAGGA	6573		
QY 1198	tctaatgtgaagaactctataaagaatttggatataatgaagaagaagtgcttcaaggtaac	1257		
Db 6574	GTTGATAGGAACACATACAGGCTGAATTCCTCATGTGGTGAGCAAGATTTTAAAGGAAT	6633		
QY 1258	tctaatcttcggagaccagatgcaaaaacattcatccctcccatcgtagccagatatgtg	1317		
Db 6634	ACTAATACCAAAAGACATGTGAAGAACTTTTCAACCCCCCAATCATTTCCAGGTTTATC	6693		
QY 1318	cgggttgcctcccaagacatgagcacaagagatagccttgaagtgagagctctgtgtgc	1377		
Db 6694	CGTGCACTTCCTAAACATGAAATCAAGTAAGTATTACACTTCGCCCTGGAACTCTTGGCTGT	6753		
QY 1378	cagatt 1383			
Db 6754	GATATT 6759			

```

1  RESULT 3
2  US-08-746-111-26
3  ; Sequence 26, Application US/08746111
4  ; Patent No. 6066778
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Ginsburg, David
9  ;
10 ; APPLICANT: Chi, Jisong
11 ;
12 ; TITLE OF INVENTION: Compositions And Methods For Screening
13 ;
14 ; TITLE OF INVENTION: Compounds For Anticoagulant Activity
15 ;
16 ; NUMBER OF SEQUENCES: 54
17 ;
18 ; CORRESPONDENCE ADDRESS:
19 ;
20 ; ADDRESSEE: Medlen & Carroll, LLP
21 ;
22 ; STREET: 220 Montgomery Street, Suite 2200
23 ;
24 ; CITY: San Francisco
25 ;
26 ; STATE: California
27 ;
28 ; COUNTRY: United States of America
29 ;
30 ; ZIP: 94104
31 ;
32 ; COMPUTER READABLE FORM:
33 ;
34 ; MEDIUM TYPE: Floppy disk
35 ;
36 ; COMPUTER: IBM PC compatible
37 ;
38 ; OPERATING SYSTEM: PC-DOS/MS-DOS
39 ;
40 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
41 ;
42 ; CURRENT APPLICATION DATA:

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,111
FILING DATE: 06-NOV-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6585 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 6..6554
US-08-746-111-4

```

```

Query Match      4.6% Score 80.6; DB 3; Length 6585;
Best Local Similarity 52.1%; Pred. No. 13e-10;
Matches 207; Conservative 0; Mismatches 184; Indels 6; Gaps 1;

```

```

QY 987 agtcaatggtctcctgagcaagccgacttcagagcaagccatcatggtctcgag 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6161 AGACTACTGGAGCCCTCCCTCCGCTGAAAGCCGCGGCGGTAAGCGCTGGCA 6220
QY 1047 cgacagtagcaaacacacacacacacacacacacacacacacacacacacac 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6221 AGCCAAGGCAAAACAAACAAAGC-----AGTGGTTACAGTCGATCTCTCAAAATCAA-6274
QY 1107 gaaatacaacagagattagagaccagatctacacagtcgaacttaactttaatgtaa 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6275 GAAGGTAAAGCGGCATGTAAGAGAGGCTGTAAAGTCTGTCTCTGAGATGTAGCTGAA 6334
QY 1167 gaatttgatgaactcaaaaataaataaataaataaataaataaataaataaataa 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6335 GACCTACAGATCCATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 6394
QY 1227 gaataatgaagaaagagtttttcagagtaacttaacttcgagaccagtgcaaaacaa 1286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6395 CTCATAGTGTGACAGATTTTGAAGAAACAGCAATACCAAGGGACATGAGACTT 6454
QY 1287 ttatccctcccatcgtgagcagatagtgaggtgtgtcccccagacatgagcagag 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6455 TTTCAACCGCCCATTTATTTTCAGATTTCGCGCATCATTCCTCAAAAACATGAAACAGAG 6514
QY 1347 gatagcctgaagagtgagcattgtgtgccaatt 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6515 CATCGCCCTTCGCGCTAGAGCTTCGCGCTGTGACATT 6551

```

RESULT 6

```

US-08-484-891-7
Sequence 7, Application US/08484891
Patent No. 5935935
GENERAL INFORMATION:
APPLICANT: Connolly, Sheila
APPLICANT: Kaleko, Michael
APPLICANT: Smith, Theodore
TITLE OF INVENTION: Adenoviral Vectors for
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gillfillian,
ADDRESSER: Cecchi, Stewart & Olstein
STREET: 6 Becker Farm Road

```

```

CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,891
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/218,335
FILING DATE: 25-MAR-1994
APPLICATION NUMBER: 08/074,920
FILING DATE: 10-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Olstein, Elliot M.
REGISTRATION NUMBER: 24,025
REFERENCE/DOCKET NUMBER: 271010-273
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4629 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna primer
FEATURE:
NAME/KEY: Factor VIII CDNA with
US-08-484-891-7

```

```

Query Match      4.5% Score 78.4; DB 2; Length 4629;
Best Local Similarity 54.1%; Pred. No. 4e-10;
Matches 210; Conservative 0; Mismatches 166; Indels 12; Gaps 2;

```

```

QY 993 ctggtctcctgagcaagcccgacttcagagcaagcccatcatggtcttcggcgagag 1052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4023 CTGCTCTCCTTCAAAAGCTCGACCTTCACCTCCAAAG-----GAGGAGTAATGCCCTGAG 4076
QY 1053 tagcaacaacacacacacacacacacacacacacacacacacacacacacacacac 1112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4077 ACCTCAGGTGAATTAATCCAAAAGAGTGGCTGCAAGTGAATTCAGAAACAAATGAAGT 4136
QY 1113 aacagaaattagacacacagatctacaacagtcgaactcaactttaattgaagatt 1172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4137 CACAGAGTAATCACTCAAGGAGTAATAATCTGCTTACACGATGTATGGAAGACTT 4196
QY 1173 tctgtagaactcaaaaataaataaataaataaataaataaataaataaataaataa 1232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4197 CCTCATCTCCACAGTCAAGATGAGCCATCAGTGAAGCTCTCTTT-----TTCAAAAAGG 4250
QY 1233 tgaagaaagtggttcaagtaacttaacttcgagaccagtgcaaaacaaatattcat 1292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4251 CAAAGTAAGGTTTTCAGGGAATCAAGACTCTTCACACTGTGGTGAAGTCTCTTGA 4310
QY 1293 cctccatcgtgagcagatagtgaggtgtgtcccccagacatgagcagagagatagc 1352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4311 CCCACCGTACTGACTCGGTACTCTTCGAATTCACCCCAAGATGTGGTGACACAGATTGC 4370
QY 1353 ctgaaagtggagctcaattgtgtccag 1380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4371 CCTGAGATGAGAGTCTGTGGCTGGAG 4398

```

RESULT 7

```

US-08-717-294-41

```

```

; Sequence 41, Application US/08717294
; Patent No. 611418
; GENERAL INFORMATION:
; APPLICANT: SEED, BRIAN
; APPLICANT: HAAS, JURGEN
; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/717,294
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/345001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-717-294-41

```

Query Match 4.5%; Score 78.4; DB 3; Length 4670;

Best Local Similarity 54.1%; Pred. No. 4e-10; Mismatches 166; Indels 12; Gaps 2;

```

Matches 210; Conservative 0; Mismatches 166; Indels 12; Gaps 2;
QY 993 ctggtctctgccaagccgaccttaagcaagcccatcatatggtcttgcggcgacg 1052
DB 4049 CTGGTCTCTTCAAAAGCTCGACTTCACCTCCAAAG-----GAGGAGTATCTCTGGAG 4102
QY 1053 tagcaaacaccacaacacgaagatggtctgagatcgatcttgaggagagaaaaaagaat 1112
DB 4103 ACCTCAGTGAATTAATCAAAAGAGTGGCTGCAAGTGGACTTCCAAAGACATGAAAGT 4162
QY 1113 aacagaatttaggaccaagatctacacagtcgaacttaacttaagttaaggttt 1172
DB 4163 CACAGAGTACTACTACAGGAGTAAATCTCTTACCAGCATGATGTGAAGAGTT 4222
QY 1173 tgtatgaactcaaaaacaaataatcttaagtgaagaccataaagaatltgaataa 1232
DB 4223 CCTCATCTCCAGACATCAAGATGCGCATGAGTGTGACTCTCTTT-----TTGAGATGG 4276
QY 1233 tgaagaagaagtgttccagggtacttaacttcgggagccagtcgcaaaacatttcac 1292
DB 4277 CAAAGTAAAGGTTTTCAGGGAATCAAGACTCTTACACACTGTGTGACTCTCTAGA 4336
QY 1293 ccctccatcgtggccagatattgctgggttgcggccagaccatggcaccagaggtacg 1352
DB 4337 CCACCGTTACTGACTCGCTACTCTTGAAATTCACCCCAAGATTGGGTGACACCAATTCG 4396

```

```

QY 1353 ctgaaggtgagctcatgtgttcgacg 1380
DB 4397 CCGAGGATGACGAGTTCTTGGGCTGCGAG 4424

```

RESULT 8

```

US-09-470-618-14
; Sequence 14, Application US/09470618
; Patent No. 6200560
; GENERAL INFORMATION:
; APPLICANT: Couto, Linda B.
; APPLICANT: Colosi, Peter C.
; TITLE OF INVENTION: Adeno-Associated Vectors for Expression of Factor VIII
; TITLE OF INVENTION: by Target Cells
; FILE REFERENCE: Avigen-04082
; CURRENT APPLICATION NUMBER: US/09/470,618
; CURRENT FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: 09/364,862
; EARLIER FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: 60/125,974
; EARLIER FILING DATE: 1999-03-24
; EARLIER APPLICATION NUMBER: 60/104,994
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 4999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-470-618-14

```

Query Match 4.5%; Score 78.4; DB 4; Length 4999;

Best Local Similarity 54.1%; Pred. No. 4.1e-10; Mismatches 166; Indels 12; Gaps 2;

```

Matches 210; Conservative 0; Mismatches 166; Indels 12; Gaps 2;
QY 993 ctggtctctgccaagccgaccttaagcaagcccatcatatggtcttgcggcgacg 1052
DB 4396 ctggtctcttcaaaagctcgacttaacctcaag-----gaggaatgcttgag 4449
QY 1053 tagcaaacaccacaacacgaagatggtctgagatcgatcttgaggagagaaaaaagaat 1112
DB 4450 acctcagtgatataatccaaagatggtctgcaagtgcacatccgaagaacatgaagt 4509
QY 1113 aacagaatttaggaccaagatctacacagtcgaacttaactttatgttaaggttt 1172
DB 4510 cacaggaataactactcagaggttaaatctctgtcttaccagcatgtatgtgaagagt 4569
QY 1173 tgtatgaactcaaaaacaaataatcttaagtgaagaccataaagaatltgaataa 1232
DB 4570 ccctcatctccagcagatcagatgccaatcagatggaactctctt-----ttcagaatgg 4623
QY 1233 tgaagaagaagtgttcaaggtacttaacttcgggagccagtcgcaaaacatttcac 1292
DB 4624 caaagtaagaagtgttccagggaatcaaacactcccttaeactctgtgtgaatccttga 4683
QY 1293 ccctccatcgtggccaagatattgctgggttgcggccagaccatggcaccagaggtacg 1352
DB 4684 ccacacgttactgactcgtactccttgaattcaccacccagagttgtgtgaccagattgc 4743
QY 1353 ctggaagtgagatcattgtgttcgag 1380
DB 4744 cctgaagatcagaggtcttggtcgcgag 4771

```

RESULT 9

```

US-09-364-862-14
; Sequence 14, Application US/09364862
; Patent No. 6221349
; GENERAL INFORMATION:
; APPLICANT: Couto, Linda B.

```

```

1  APPLICANT: Colosi, Peter C.
2  TITLE OF INVENTION: ADEMO-ASSOCIATED VECTORS FOR EXPRESSION OF FACTOR VIII
3  TITLE OF INVENTION: BY TARGET
4  TITLE OF INVENTION: CELLS
5  FILE REFERENCE: AVIGEN-03743
6  CURRENT APPLICATION NUMBER: US/09/364,862
7  CURRENT FILING DATE: 1999-07-30
8  EARLIER APPLICATION NUMBER: 60/125,974
9  EARLIER FILING DATE: 1999-03-24
10 EARLIER APPLICATION NUMBER: 60/104,994
11 EARLIER FILING DATE: 1998-10-20
12 NUMBER OF SEQ ID NOS: 14
13 SOFTWARE: PatentIn Ver. 2.0
14 SEQ ID NO 14
15 LENGTH: 4999
16 TYPE: DNA
17 ORGANISM: Artificial Sequence
18 FEATURE:
19 FEATURE INFORMATION: Description of Artificial Sequence: Synthetic
20 US-09-364-862-14

```


Search completed: July 9, 2002, 01:07:08
Job time: 5575 sec

```

Db 6717 ACCTGAGTGAATAATCCAAAGAGTGGCTGCAAGTGAAGTCTCCAGAAACATGAAGT 6776
QY 1113 aacaggaaattgagccacagagatctacacagtgcaacttaacttattatgtaagattc 1172
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6777 CACAGGAGTAACCTACTCAGGAGTAATAATCTCTCTTACACGACATGATGGAAGGATT 6836
QY 1173 tggatgaacttcaaaaacaaataatctaaagtggaagacctataaagaaattgtaataa 1232
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6837 CCTCATCTCCAGCAGTCAAGATGCGCATCAAGTGAAGTCTCTTTT-----TTCGAATGG 6890
QY 1233 tgaagaaaagggtgttcaagggttaacttaacttcgggaccagtgcaaaacaaattcat 1292
    ||| ||||| ||||| ||||| ||| | | | | | | | | | | | | | | | |
Db 6891 CAAAGTAAGGTTTTTTCAGGGAATCAAGACTCCTTCACACCTGCTGGTGAACCTCTAGA 6950
QY 1293 cccctccatcgtggccagatgtgcccagagatgtcccccagacatgycacacagagatagc 1352
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6951 CCCACCGTTACTGACTGCTACTCTTCGAATTCACCCCCAGAGTTGGGTGCAACGATTCG 7010
QY 1353 ctgaaggtggagctcatgtgttgccag 1380
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7011 CCTGAGATGAGGCTCTGGGCTGCGAG 7038

```

```

RESULT 15
5171844-1
; Patent No. 5171844
; APPLICANT: VAN OUYEN, ALBERT J. J.; PANNEKOEK, HANS; VERBEET,
; MARTINUS P.; VAN LEEN, ROBERT W.
; TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY
; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS
; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/205,226
; FILING DATE: 10-JUN-1988
; SEQ ID NO:1:
; LENGTH: 8241
5171844-1

```

```

Query Match 4.5%; Score 78.4; DB 6; Length 8241;
Best Local Similarity 54.1%; Pred. No. 5.1e-10;
Matches 210; Conservative 0; Mismatches 166; Indels 12; Gaps 2;

QY 993 ctggtctcctgccaagccgacttcaggaacagccacatcagtggtcttcggcgagag 1052
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6669 ctggtctcctccaagaagctgacttcacctccaag-----gaggaatgactcgag 6722
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1053 tagcaacaacacacacagagatggtcgtgagatcgatttgggggagaaaaaagaat 1112
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6723 acctcaggtgaataatccaaaagatggtcgaagtggacttcagaagaacaatgaagt 6782
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1113 aacaggaattagacacacagatctacacagtcgaactcaacttattatgtaagattt 1172
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6783 cacaggaatgaactacacagagtaaaatctcgtcttacacagatglatgtgaagagtt 6842
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1173 tggatgaacttcaaaaacaaataatctaaagtggaagacctataaagaaattgtaataa 1232
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6843 cctcatctccagcagtcagaatgcatcagtgactctctt-----ttcagaatgg 6896
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1233 tgaagaaaagggtgttcaagggttaacttaacttcgggaccagtgcaaaacaaattcat 1292
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6897 caaagtaaaagtttccaagggaatacaagactccttcacacacttggtgaactctaga 6956
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1293 cctcccatcgtggccagatgtgcggtgtgccccagacatgycacacagagatagc 1352
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6957 cccacaggttactgactcgtcactcgaatcccccacagagltgggtgacacagattgc 7016
    ||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1353 ctgaaggtggagctcatgtgttgccag 1380
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7017 cctgagatggaaggtctcgtggtcgag 7044

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 01:07:13 : Search time 80.25 Seconds

(without alignments)
811.081 Million cell updates/sec

Title: US-09-691-344A-4

Perfect score: 3052
Sequence: 1 MGFAGAGRLRPVAPRSSAE.....YDNEKENTQKLDITSOMAG 586

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3052	100.0	586	22 AAU00629	Novel human protei
2	2818	92.3	539	22 AAU00630	Novel human protei
3	2812	92.1	715	22 AAU00670	Human TANGO 229 po
4	2557	83.8	487	22 AAU00628	Polypeptide isolat
5	1301.5	42.6	503	21 AAB19126	Human human protei
6	903.5	29.6	583	22 AAG75450	Human colon cancer
7	786.5	25.8	669	21 AAY70539	Human Factor 8 Hom
8	353.5	11.6	889	21 AAB24216	Soluble neuropilin1
9	351.5	11.5	925	20 AAW96308	Neuropilin2. Rat
10	349	11.4	439	22 AAB64627	Human secreted pro
11	349	11.4	555	20 AAY06320	Human soluble neur

12	349	11.4	909	20 AAW96254	Mouse semaphorin r
13	349	11.4	926	20 AAW96255	Human semaphorin r
14	349	11.4	931	20 AAY06318	Human neuropilin-2
15	349	11.4	931	20 AAY14563	Human VEGF(165)/R/N
16	349	11.4	931	20 AAY23248	Human NP-2 protein
17	345.5	11.3	487	22 AAB64570	Human secreted pro
18	343.5	11.3	901	20 AAW96256	Mouse semaphorin r
19	343.5	11.3	906	20 AAW96257	Mouse semaphorin r
20	343.5	11.3	909	20 AAW96249	Rat semaphorin rec
21	343.5	11.3	909	20 AAW96250	Mouse semaphorin r
22	343.5	11.3	909	21 AAB24214	Mouse soluble neur
23	343.5	11.3	914	20 AAW96251	Mouse semaphorin r
24	343.5	11.3	926	20 AAW96252	Mouse semaphorin r
25	343.5	11.3	931	20 AAW96253	Mouse semaphorin r
26	343.5	11.3	957	22 AAW25770	Human protein sequ
27	341	11.2	439	22 AAB64628	Human secreted pro
28	335	11.0	538	22 AAU02950	Angiotensin conver
29	335	11.0	600	22 AAU02948	Angiotensin conver
30	335	11.0	644	20 AAY06319	Human soluble neur
31	335	11.0	840	22 AAU02949	Angiotensin conver
32	335	11.0	856	22 AAB62478	Human NP-1 recepto
33	335	11.0	923	20 AAY06317	Human neuropilin-1
34	335	11.0	923	20 AAY23247	Human VEGF165R/NP-
35	335	11.0	923	20 AAW96246	Human semaphorin r
36	335	11.0	924	22 AAB62476	Human neuropilin-1
37	331	10.8	922	20 AAW96309	Neuropilin. Ratn
38	329	10.8	892	21 AAB24215	Soluble neuropilin
39	329	10.8	923	20 AAW96248	Mouse semaphorin r
40	325.5	10.7	923	21 AAB24213	Mouse soluble neur
41	325.5	10.7	921	20 AAW96247	Rat semaphorin rec
42	318	10.5	2224	20 AAY49564	Human lipoprotein
43	318	10.4	2183	21 AAB03533	Murine factor V SE
44	317	10.4	2224	17 AAW04254	Human Factor V. H
45	315	10.3	159	22 AAB71907	Human FV C2 domain

ALIGNMENTS

RESULT	ID	AAU00629	standard; Protein; 586 AA.
1	AAU00629		
XX	AAU00629		
XX	29-AUG-2001	(first entry)	
XX			
DE	Novel human protein (NHP) sequence #2.		
KW	Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;		
KW	obesity; high blood pressure; connective tissue disorder; infertility;		
KW	NHP-mediated pathway.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200129219-A1.		
XX			
PD	26-APR-2001.		
XX			
PF	08-OCT-2000; 2000WO-US28798.		
XX			
PR	19-OCT-1999; 99US-0160285.		
PR	18-FEB-2000; 2000US-0183583.		
XX			
PA	(LEXI-) LEXICON GENETICS INC.		
XX			
PI	Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;		
XX	WPI: 2001-290917/30.		
DR	N-PSDB: AAS00614.		
XX			
PT	Novel nucleic acid encoding human CUB-domain containing protein, useful		
PT	for drug screening, diagnosis and treatment of physiological disorders		

PT	or diseases -
XX	
PS	Claim 2; Page 27-28; 33pp; English.
XX	
CC	The sequence represents a novel human protein (NHP) containing a CUB
CC	domain (an extracellular domain). CUB proteins have been associated with
CC	regulating development, modulating cellular processes and preventing
CC	infectious disease. NHP nucleotide sequences are useful for gene therapy
CC	of physiological disorders or diseases. NHP oligonucleotides are useful
CC	as hybridisation probes for screening libraries and assessing gene
CC	patterns. NHP nucleotide sequences are useful for detecting mutant or
CC	inappropriately expressed NHPs (for example, those proteins associated
CC	with obesity, high blood pressure, connective tissue disorders and
CC	infertility) for the diagnosis of a disease. The polynucleotides may also
CC	be used in screening for drugs effective in the treatment of symptomatic
CC	or phenotypic manifestations of perturbing the normal function of NHP in
CC	the body. Nucleotide constructs encoding NHP products are used to
CC	genetically engineer host cells to express such products in vivo. These
CC	host cells allow for the identification of compounds that bind to NHP
CC	receptors or trigger NHP-mediated pathways.
XX	
XX	
Sequence	586 AA;

Query Match	100.0%;	Score 3052;	DB 22;	Length 586;
Best Local Similarity	100.0%;	Pred. No. 5.7e-265;		
Matches 586;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MGGAGORLRLPVPAPRRSSSEEARPGCOLRIGIRRGAELEKLAPSCVMPGARGGALAR	60
Db	1	mgfagqrllrpppapsseeeatppqqlrlgllrrgeaelaklapsevmpgarrggatar	60
QY	61	AAGRGUALLLVASAPLRLOAEELDGDGGHLYVYODSGWTMSKNYBGPYPNHVTVEKTYT	120
Db	61	aagrgllalllvasaprlrlgaeelgdggghlyvvyqdsagntsmnypgypnhltvecklt	120
QY	121	VPEKRLILRLGDDIESQFCASDYLFTSSSDQYGYGSGMTVPKELLMTSEVTRFE	180
Db	121	vpekrlilrlrgdlleeqcaesdyllftssdsdqygycgsmvpekellmtsevtvrf	180
QY	181	SGSHISGREFELLYVASSPHDLITLCERASHLYLTXEYSKRCPCAGCDVAGDISGMNVGY	240
Db	181	sgshisgrefillyvassdphdlitclerashlyltxeyskrcpcagcdvagdsgnmvgy	240
QY	241	RDVSLICKAKAIHAGITADELGGQISVLORRKISREYELLANGVLSRDGSLSPKRLPFNS	300
Db	241	rdvsllickaaihagitadeelggqisvlorrkisreyellangvlsrdgslskskrllfns	300
QY	301	GCGRSLSEFPDQOITRASSSMQSVNESGDQVHMSFGQARLDQGPSWASGDSSNNHKPREW	360
Db	301	gcgrslsefpdqoitraasssmqsvnesgdqvhwmsfgqarlldqgpswasgdssnnhkprew	360
QY	361	LEIDJGEKKKINGITTTSTQSNFNFYKSPVMYFKNNNSKMYTGYVNNNEKTFQONS	420
Db	361	leidjgekkkigtlttstqsnfnfykspvmfknnskmytgyvnnnektfqons	420
QY	421	NFRDPVONNFPIPIYARVYVVPQTMHORIALKVELIGCOITFGNSLSLWKRSTOSTSVS	480
Db	421	nfrdpvonnfpipiyarvyvvpqtmhorialkveligcoitfgnslslwkrstostsvs	480
QY	481	TKREDETTIRPISEESTGTGINTTVAIPLVLLVLFVAFAGMIGFAAFRRKKKKGGSPYGA	540
Db	481	tkredeettirpiseestgtginttvaipvlvllvlfvafagmigfaafrrkkkkggspysa	540
QY	541	EAKOTCQKQIKYPPARKQSAEFTITSDNEKEMTQKDLITSMAG	586
Db	541	eakotcwkqikypparkqsaeftitstnekemtkqldlitsmag	586

RESULT	2
AAU00630	
ID	AAU00630 standard; Protein; 539 AA.
XX	

AC AAU00630;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #3.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PW WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00615.
DR

Claim 5; Page 29-30; 33pp; English.

The sequence represents a novel human protein (NHP) containing a CUB domain (an extracellular domain). CUB proteins have been associated with regulating development, modulating cellular processes and preventing infectious disease. NHP nucleotide sequences are useful for gene therapy of physiological disorders or diseases. NHP oligonucleotides are useful as hybridisation probes for screening libraries and assessing gene patterns. NHP nucleotide sequences are useful for detecting mutant or inappropriately expressed NHPs (for example, those proteins associated with obesity, high blood pressure, connective tissue disorders and infertility) for the diagnosis of a disease. The polynucleotides may also be used in screening for drugs effective in the treatment of symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Nucleotide constructs encoding NHP products are used to genetically engineer host cells to express such products *in vivo*. These cells allow for the identification of compounds that bind to NHP receptors or trigger NHP-mediated pathways.

SQ Sequence 539 AA;

Query Match	92.3%;	Score 2818;	DB 22;	Length 539;
Best Local Similarity	100.0%;	Pred. No. 5.2e-244;		
Matches 539;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	48	MYPGRGGALARAAGRLALLLAASAPRLAAEELGDCGHHVTVQDSGWTMSKPYG	107
Db	1	mvpa9ggagalaaraagrlalllaasaprlrlgaaei9dg9ghnlvtyqds9tmskrypp	60
QY	108	TYPNHWCEKRTTPPKGRLLRLRGDLDIESQTCASDYLLFTSSSDQYGYCGSMTPKE	167
Db	61	lypnhtvekekltyrpkgrllrlrlrgddiesqtcasdyllftssddqygycgsmtpke	120
QY	168	LLLNTSEVTVAFESGSHISGRGLLTVAASSDHPDLITCLEPASHYLTKEYSKFCPACRD	227
Db	121	lllntsevtvrfesgshisgrgfllyaaashdpdlitclerashylkkeyskfcpagcrd	180
QY	228	WAGDSGMDWCVGRDTSLLCKAAHAHAGLIDELGGQISVQRKCIKRYEGLANGVSRD	287
Db	181	vagdsqmdwcvgyrdtsllckaaahagladelggqislvqrkqklsryeglanguvrd	240

SEQ ID NO: 4
Database: A-Geneseg-032802
Accession: AAU00670

```
QY 288 GSLSDKRFLETSNGCSRSLSFEPDGOIRASSWSQSVNESCDOVHMSPGQARLQDQGPSWA 347
    |||||||
Db 241 gslsdkrflfetsngcsrslsfepdqgdirsswsgvnesgdqvhmspgqarlgddqgpswa 300
QY 348 SGDSNNHPRPRLMELIDGEEKKRTGIRTTGSTQSNFNFYKSFYMNFKNNNSKKMTYKG 407
    |||||||
Db 301 sgdsnnhprwleldlgckkrtgirtgstqsnfnfykxfymnfknnskwtkyg 360
QY 408 IYNEEKYVQGSNFRDPVQNNFPIPIVARYRVVPQTHQRIALKVELIGCQITQGNDS 467
    |||||||
Db 361 iyneekvqgsnfrdpvqnnfipipivaryrvvpqthqriaikveligcqitqgnds 420
QY 468 IYWRKTSOSTSVSTKKEDETIRPISEETSGINITVAIPVLVLYLVFAGMGIFAFAF 527
    |||||||
Db 421 iwrktsqstsvstkkedeltirpiseetstginittvaipvlvlylvfagmgifaaf 480
QY 528 RKKKKKSPYGSAAQKTDCKWQIKYPPARHQAETTSYDNKEKMTOKLIDITSDMA 586
    |||||||
Db 481 rkkkkgspsygsaaqktdcwkwikypfarhgsaetflsydnkemtkgldlitsdma 539
```

RESULT 3

AAU00670
ID AAU00670 standard; Protein: 715 AA.

AC AAU00670;

DT 07-SEP-2001 (first entry)

DE Human TANGO 229 polypeptide.

Human: TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung; skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte; bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer; anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis; attention deficit disorder; Crohn's disease; gastroenteritis; goitre; hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism; muscular dystrophy; immuno-competence; vertebrate; blood; serum.

OS Homo sapiens.

Key Location/Qualifiers

FT Peptide 1..34 /note= "Signal peptide"

FT Domain 35..455 /note= "Extracellular domain"

FT Protein 35..715 /note= "Mature human TANGO 229"

FT Domain 456..480 /note= "Transmembrane domain"

FT Domain 481..715 /note= "Cytoplasmic domain"

PN WO200129088-A1.

PD 26-APR-2001.

PF 23-JUN-2000; 2000WO-US17386.

PR 19-OCT-1999; 99US-0420707.

PA (MILL-) MILLENNIUM PHARM INC.

PI Mackay CR, Myers PS, Kilst SJ, Fraser CC, Leiby KR.

DR WPI: 2001-308477/32.

DR N-PSDB: AAS00660.

PT New isolated nucleic acid molecule for diagnosis, prevention, and therapy of human and other animal disorder, or as modulating agent for regulating cellular processes

XX Claim 8; Fig 1; 263pp; English.

PS The sequence represents human TANGO 229 polypeptide. This protein and
CC similar others exhibit the ability to affect growth, proliferation,
CC survival, differentiation, activity, morphology, or movement/migration
CC of, e.g. T cells and cells of the heart, liver, pancreas, placenta,
CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
CC blood leukocyte, bone marrow or thymus tissue. They can be used as
CC modulating agents for regulating cellular processes, thus, the proteins
CC and their associated nucleic acids can be used to prognosticate, prevent,
CC diagnose, or treat disorders associated with physiological processes.
CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,
CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
CC deficit disorder, Crohn's disease, gastroenteritis, goitre,
CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
CC muscular dystrophy. Antibodies to disorders such as these can be made by
CC providing a polypeptide of the invention to an immuno-competent
CC vertebrate and harvesting blood or serum from the vertebrate.

XX Sequence 715 AA:

Query Match 92.1%; Score 2812; DB 22; Length 715;
Best Local Similarity 100.0%; Pred. No. 2,8e+243;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 MYPGARGGALARAANGRLALLAVSAPRLQAEELDCGCHLYTVQDSCSTMSTKPYG 107
Db 1 mypgarggalaraaegrllallavsaprlqaeelgdgcghlytvqdsqtmstsknypg 60

QY 108 TYPNHTVEKTTTPKGRRLIRLGDLESGTQSDVLTLTSSSDQGPYCGSMTVPKE 167
Db 61 typnhtveekttytpkgrllirldesgtcsdyltltssdqgpycgsmtyvke 120

QY 168 ILLNLTSEVTVRPESGSHISGRGFLTTVASDHPDLITLERASHLYLKEYSKFCPPAGCD 227
Db 121 illnltsevtvresgshisgrgflttvasdhpdlitlerashlylkeykfcppagrd 180

QY 228 VAGDISGNWDTGRTSLCKRAHAGIITADELGOIYVLRKGSIRREGILANGVLSRD 287
Db 181 vagdisgnwtdyrtslckraahagiitadelgqisvlgrkgsirregillangvisrd 240

QY 288 GSLSDKRFLETSNGCSRSLSFEPDGOIRASSWSQSVNESCDOVHMSPGQARLQDQGPSWA 347
Db 241 gslsdkrflfetsngcsrslsfepdqgdirsswsgvnesgdqvhmspgqarlgddqgpswa 300

QY 348 SGDSNNHPRPRLMELIDGEEKKRTGIRTTGSTQSNFNFYKSFYMNFKNNNSKKMTYKG 407
Db 301 sgdsnnhprwleldlgckkrtgirtgstqsnfnfykxfymnfknnskwtkyg 360

QY 408 IYNEEKYVQGSNFRDPVQNNFPIPIVARYRVVPQTHQRIALKVELIGCQITQGNDS 467
Db 361 iyneekvqgsnfrdpvqnnfipipivaryrvvpqthqriaikveligcqitqgnds 420

QY 468 IYWRKTSOSTSVSTKKEDETIRPISEETSGINITVAIPVLVLYLVFAGMGIFAFAF 527
Db 421 iwrktsqstsvstkkedeltirpiseetstginittvaipvlvlylvfagmgifaaf 480

QY 528 RKKKKKSPYGSAAQKTDCKWQIKYPPARHQAETTSYDNKEKMTOKLIDITSDMA 585
Db 481 rkkkkgspsygsaaqktdcwkwikypfarhgsaetflsydnkemtkgldlitsdma 538

RESULT 4

AAU00628
ID AAU00628 standard; Protein: 487 AA.

AC AAU00628;

DT 29-AUG-2001 (first entry)

XX

DE Novel human protein (NHP) sequence #1.
XX
XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
XX WPI: 2001-290917/30.
DR N-PSDB: AAS00613.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
XX
PS Claim 4; Page 26; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
CC
XX
XX
SQ Sequence 487 AA:

Query Match 83.8%; Score 2557; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 1,2e-220;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 MTSKNVPCGYPMHTVCEKITYPKGRRLRLRGDLDIESQTCASDYLLFTSSDDGYPC 159
Db 1 mtsknpvcgyphvcektlvpkgrllrlrgdlldiesqtcasyllftssddgyppc 60

QY 160 GGMTVPKELLMTSEVTFEFGSHISGRFLTYAASSHPDLITCLERASHYLTPEVSK 219
Db 61 ggmtyvpkellmtsevtvfeegshisgrflltyaasshpdlltclerashyltpevsk 120

QY 220 FCPACCRDVAAGDISGNMVDGYRDTSLCKAAIHAGIINDELGGQISVLRKQISRYEGIL 279
Db 121 fcpagcrdvagdisgnmvdgyrdtsslckaaiahagiindelggqisvlgkqisryegil 180

QY 280 ANGVLSRDSGLSDKRFRTSNCCSRSLSPFPGQIRASSWQSVNBSGDVHWSGQARL 339
Db 181 angvlsrdsghlsdkrftrsnccsrslspfpqgqirasswqsvnbsgdvhwspgqarl 240

QY 340 OQOGSMASGSSNNHKKPREMLEIDLGEKKRTGTGRTGTSQSNFVYKSVVMFKNN 399
Db 241 qogsgmasgssnnhkkpremlleidlgekkrtgtgtrtgtsqsnfvvyksvvmfknn 300

QY 400 SKMKTGKIVNNEEKVFGNSNFRDPVONNFIPIVARYVRVPOTWQRIALKEVILGC 459
Db 301 skmktgkivnneekvfgnsnfrdpvonnfiipivaryvrvpotwqrialkvelilgc 360

QY 460 QITGNDLSLWKRISQTSVSTKKEDEITTRIPSEETSGINTTVAIPVLVLVFA 519
Db 361 qitgndslwkrtsqtsvstkkedelttripseetstginttvaipvlvlvfa 420

QY 520 GNGIFAAFRKKKKSPYGSAAQKTDCKQIKYPAFHQSNFPISTNDKEMTKQKDL 579
Db 421 gngifaafrkkkspygsaaqktdckwikypafhqsnfpiistndkemtkqkdl 480

QY 580 ITSDMAG 586
Db 481 itsdmag 487

RESULT 5
AAB19126
ID AAB19126 standard; Protein: 503 AA.
AC AAB19126;
XX
XX 19-FEB-2001 (first entry)
DE Polypeptide isolated from lymph node stromal cells of fsn -/- mice.
XX
XX Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW immune system disorder; cancer; viral infection; HIV infection;
KW blood vessel growth; tumour necrosis factor disorder; arthritis;
KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
XX cardiac failure.
OS Mus sp.
XX
XX WO200058463-A1.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 18-FEB-2000; 2000WO-NZ00015.
PP
XX
XX 25-MAR-1999; 99US-0276268.
PR
XX 26-AUG-1999; 99US-0383586.
PR
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA
XX
XX Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI
XX
XX Murison JG;
XX
XX WPI: 2000-664924/64.
DR
XX N-PSDB: AAA96736.
XX
PT Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
PT useful for modulating growth of blood cells, for treating inflammatory
PT and tumour necrosis factor-mediated disorders, cancer and viral
PT disorders -
XX
XX
XX
PS Claim 1; Page 68-69; 75pp; English.
XX
XX The present sequence represents a polypeptide sequence which is
XX isolated from lymph node stromal cells of fsn -/- mice. The
XX polynucleotides and their polypeptides are useful for treating an
XX inflammatory disorder, disorder of immune system and cancer selected
XX from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
XX viral disorder, in particular HIV infection and for modulating the
XX growth of blood vessels. The polypeptides are useful for treating a
XX tumour necrosis factor (TNF) mediated disorder, such as those selected
XX from arthritis, inflammatory bowel disease and cardiac failure and a
XX fibroblast growth factor-mediated disorder. It is also useful in assays
XX to determine biological activity, to raise antibodies, to isolate
XX corresponding ligands or receptors, to quantify levels of protein or
XX cognate corresponding ligand or receptors, as anti-inflammatory agents,

QY 238 DGYRDTSLCKAIIHAGIIADELGQISVLQKKGISRYEGILANGVLSRDGSLSDKRFLE 297

Query Match 11.6%; Score 353.5; DB 21; Length 889;

KW Cytostatic; immunosuppressive; neutrophic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antituber;
 KW vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.

PN WO20077197-A1.

PD 21-DEC-2000.

PF 01-JUN-2000; 2000WO-US14934.

PR 11-JUN-1999; 99US-0138599.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM, Komatsoulis GA;

DR WPI; 2001-032312/04.

PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -

PS Disclosure; Page 524-526; 558pp; English.

XX The invention relates to the isolation of genes AAF32757-F32803 encoding
 CC the human secreted proteins AAB64549-B64594. The sequence is a search
 CC result from a BLASTX homology search. The genes and proteins are useful
 CC for preventing, ameliorating or treating medical conditions, e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.

XX Sequence 439 AA;

Query Match 11.4%; Score 349; DB 22; Length 439;

Best Local Similarity 25.3%; Pred. No. 1.9e-22;

Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

QY 88 GCHLVYQDSGTMKSNKYGTYPNHTVECKITVPR-GKRLILRG-DLIDISQICADSY 145
 DB 28 GCGRIKAAAGYITSPYQDYPQDYSPQNEWYAYAPENQKIVLINFPHFELEKHCXDF 87
 QY 146 LTF---TSSDQGYGCGSMVTPKELLNLTSEVVRFECSHGIRGFLTY-----AS 196
 DB 88 IETIRGDSADLLGHKGNL-APTLLSSGSMLYIKFTSDARGAGFSLYELFKGYS 146
 QY 197 SD-----HPDLITC---LERASHYKTEYSKF----- 220
 DB 147 EDCSKNFTSPNGTLESPPFKYPHNLDOCTFLIKPMELLGFLIDLENDPQVYGE 206
 QY 221 -CPACRYVAGDIS--GMMVDCYRDTSLICKAATAGIADLGGOISYLQKRGISRIEG 277
 DB 207 DCKYGLDWGDIPIHVPLIKYCGTKPSLRSETGLISLTFHCDMAVAK-----DG 259
 QY 278 ILANGVLSRDSGLSKRFLFTSNGCSRSLSFE---PDGQIRASSSMQSVNCSGQVHMS 333
 DB 260 fSaryylvnhgeplnfq-----cnvplgmesgrlneqdsassty-----sdgrwt 305

QY 334 PQCARLADQGPSPWAGSDSSNNHKPREMLEIDLGEKKKITGIRTG--STQSNFNFYVKSF 391
 DB 306 PQGSLHNGDDNGWTPNLSN-----KEYLQVDLIRFLMTATATGALSRENGYVYKSY 361
 QY 392 VMEFRNNNSKKMTYGIYVNEEKVFQGSNFRPDYQNNFIPPYARYVRVPPQWQHRIA 451
 DB 362 KLEVSTNGEDMNVYHGXNH--KVFGANNDADEVLNKHAPLIRLFRVIRPQLTWSGIA 419

QY 452 LKVELIGCOIT 462
 DB 420 IRLTelfgrvt 430

RESULT 11

ID AAY06320 standard; Protein: 555 AA.

AC AAY06320;

DT 06-SEP-1999 (first entry)

DE Human soluble neuropilin-2.

XX Neuropilin-2; human; vascular endothelial growth factor receptor;
 KW VEGFR15R; inhibitor; metastasis; angiogenesis; inflammation;
 KW arthritis; diabetic retinopathy; cyostatic; antiinflammatory;
 KW angiogenic; antidiabetic; ophthalmological; antiarthritic;
 KW soluble.

OS Homo sapiens.

OS Synthetic.

PN WO9929858-A1.

PD 17-JUN-1999.

PF 09-DEC-1998; 98WO-US26138.

PR 09-SEP-1998; 98US-0099615.

PR 09-DEC-1997; 97US-0069155.

PR 12-DEC-1997; 97US-0069687.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI Gagnon ML, Klagsbrun M, Soker S;

DR WPI; 1999-394975/33.

DR N-PSDB; AAX39154.

PS Soluble human neuropilins and related polynucleotides

XX Claim 6; Page 90-91; 98pp; English.
 CC The present sequence represents a soluble, truncated form of human
 CC neuropilin-2, a novel receptor of vascular endothelial growth
 CC factor (VEGF). The soluble neuropilin-2 deviates from the
 CC full-length protein (see AAY06318) between the b2 and c domains
 CC after amino acid 547, at the position of an exon-exon boundary,
 CC and includes 8 extra C-terminal amino acids. Soluble neuropilins
 CC are capable of binding to VEGFR165 or any form of VEGF that contains
 CC exon 7 and are useful for inhibiting VEGF interaction not only
 CC with neuropilins but also with KDR/Flk-1 and Flt1. They may also
 CC act as dominant negative receptors when expressed in cells by
 CC dimerizing with intact neuropilin receptors. The soluble
 CC neuropilins are useful for treatment of diseases or disorders
 CC associated with VEGF, e.g. metastasis, inappropriate angiogenesis,
 CC chronic inflammation, diabetic retinopathy and arthritis. Solid
 CC tumors can also be treated.

XX Sequence 555 AA;

Query Match 11.4%; Score 349; DB 20; Length 555;

Tue Jul 9 08:27:35 2002

us-09-691-344a-4.rag

COPY 9 A-5'
Page 12

8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 01:12:13 ; Search time 66.82 seconds
(without alignments)
214.208 Million cell updates/sec

Title: US-09-691-344A-4
Perfect score: 3052
Sequence: 1 MGFAGGRLRPVPAFRSSAE.....YDNEKEMTKLDTISWAG 586

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCOTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	349	11.4	909	3	US-08-936-135-18 Sequence 18, Appl
2	349	11.4	926	3	US-08-936-135-20 Sequence 20, Appl
3	343.5	11.3	901	3	US-08-936-135-22 Sequence 22, Appl
4	343.5	11.3	906	3	US-08-936-135-24 Sequence 24, Appl
5	343.5	11.3	909	3	US-08-936-135-8 Sequence 8, Appl
6	343.5	11.3	909	3	US-08-936-135-10 Sequence 10, Appl
7	343.5	11.3	926	3	US-08-936-135-12 Sequence 12, Appl
8	343.5	11.3	914	3	US-08-936-135-14 Sequence 14, Appl
9	343.5	11.3	931	3	US-08-936-135-16 Sequence 16, Appl
10	329	10.8	923	3	US-08-936-135-6 Sequence 6, Appl
11	321	10.5	218	1	US-07-607-538C-4 Sequence 4, Appl
12	321	10.5	218	1	US-08-162-402B-4 Sequence 4, Appl
13	318	10.4	2183	3	US-08-746-111-5 Sequence 5, Appl
14	311	10.2	160	2	US-08-162-402B-14 Sequence 14, Appl
15	306.5	10.0	2343	4	US-09-324-867-2 Sequence 2, Appl
16	299	9.8	217	1	US-07-607-538C-3 Sequence 3, Appl
17	299	9.8	217	2	US-08-162-402B-3 Sequence 3, Appl
18	299	9.8	218	1	US-07-607-538C-2 Sequence 2, Appl
19	299	9.8	218	2	US-08-162-402B-6 Sequence 2, Appl
20	299	9.8	387	2	US-08-162-402B-6 Sequence 6, Appl
21	299	9.8	465	2	US-08-162-402B-8 Sequence 8, Appl
22	293	9.6	159	2	US-08-162-402B-12 Sequence 12, Appl
23	289.5	9.5	2351	4	5422260-1 Patent No. 5422260
24	285.5	9.4	1438	4	US-09-209-916-1 Sequence 1, Appl
25	285.5	9.4	1471	1	US-08-683-839B-3 Sequence 3, Appl
26	285.5	9.4	1661	2	US-08-882-083-2 Sequence 2, Appl
27	285.5	9.4	1661	2	US-08-558-107-2 Sequence 2, Appl

28	285.5	9.4	1661	4	US-09-243-539-2 Sequence 2, Appl
29	285.5	9.4	2332	1	US-07-864-004B-4 Sequence 4, Appl
30	285.5	9.4	2332	1	US-08-251-937A-4 Sequence 4, Appl
31	285.5	9.4	2332	1	US-08-212-133A-2 Sequence 2, Appl
32	285.5	9.4	2332	1	US-08-276-594A-2 Sequence 2, Appl
33	285.5	9.4	2332	1	US-08-474-503-2 Sequence 2, Appl
34	285.5	9.4	2332	2	US-08-670-707A-2 Sequence 2, Appl
35	285.5	9.4	2332	4	US-09-037-601-2 Sequence 2, Appl
36	285.5	9.4	2332	4	US-09-324-867-3 Sequence 4, Appl
37	285.5	9.4	2332	5	PCT-US93-03375-4 Sequence 4, Appl
38	285.5	9.4	2332	5	PCT-US94-13200-2 Sequence 2, Appl
39	285.5	9.4	2351	1	US-08-121-202-2 Sequence 2, Appl
40	285.5	9.4	2351	1	US-08-366-851A-2 Sequence 2, Appl
41	285.5	9.4	2351	6	5171844-2 Patent No. 5171844
42	284.5	9.3	218	1	US-07-607-538C-5 Sequence 5, Appl
43	284.5	9.3	218	2	US-08-162-402B-5 Sequence 5, Appl
44	279.5	9.2	160	2	US-08-162-402B-15 Sequence 15, Appl
45	276.5	9.1	2304	4	US-09-324-867-4 Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-08-936-135-18
; Sequence 18, Application US/08936135
; Patent No. 6054293
;
GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAM GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-936-135-18

Query Match 11.4%; Score 349; DB 3; Length 909;
Best Local Similarity 25.3%; Pred. No. 8.6e-24;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

OY 88 CGHLYVODSGTMTSKYPTGYNHNYCEKTIIVPK-GKRLILRLG-DDIDESQTCASDY 145
DB 28 CGGRLNSKDGAYTSPGYDPDYSHONCEWIVYAPENOKIVLNFPHFEIKHCKYDF 87
```

```
Query Match      11.4% Score 349; DB 3; Length 926;  
Best Local Similarity   25.3%; Pred.No.8.9e-24;  
Matches 109; Conservative    72; Mismatches 166; Indels     84; Gaps    16;
```

DY
CGHLVTVODSGTMTSKNVPQTYPNHTVCERTITVPK-GKRLILRLG-DLIDIESOTCASDY 145

DB
CGRNLNSKDAGYTITSFGYPQDYPSSHONCEMIYYAPENPCKIVLNFPHPREIKHDOKYDF 87

DY
146 LAL-----TSDDPYGYCCSMTPVKRLLINTSEVTYRFPSGSIHISRGFLITY-----AS 196

DB
IELRDGDSESADILLGNHCGNT-APTTISSGLMYLKFTSDIAROGACGFSLRYEIFTGTCS 146

DY
197 SD-----HPDLITE-----LERASHLYTKTESKF----- 220

DB
EDCSKNFSPNGTIEESPGPEKYPHNLDCTFTTLAKPKMEIILOPLIFLEHDPLOYEGEC 206

DY
221 -CPAGRDAVGDS--GNMVDGYRODTSLCLCAAIAHGIIADELGCIOISVLQRKGISRREG 277

DB
207 DCKYMDLTDWIGIPHYGPLLGRCYCKRPSELRSSTYGILTLPFHDMVAK-----DG 259

DY
278 ILANGVLSRDGSTLRFRFLTSCNRSLSPFE----PDGOIRASSSWSOVNESGDQVHMS 333

DB
260 FSARYLVLAHQEPLENQ-----CNVRPIGMESGRANQAISAISTY-----SDGKRKT 305

DY
334 PGCAQLQDOGPPSWASDSSNNHKPREMLEIDIGEKKKITGITRTTG-STQSFNFYKSF 391

DB
306 POOSRIJHGGDNQWTMLDNS---KEYIQVDLRFULTMLTAITQGAISREFONGYGVXSY 361

DY
392 VNMFKNNSKMVTKGIYVNEKEVRGANSNFDPQNNPFIPIVARVAVPOOTHOLA 451

DB
362 KLEVSTINGEDMAVYRRGNKH-KVFQANDATEEVYLKLHADLFETVRIKPQTWHSGIA 419

DY
452 LAVELICQIT 462

DB
420 LRLELFQCXYT 430

RESULT 3
US-08-936-135-22
Sequence 22, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936.135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 901 amino acids

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-8

Query Match 11.3%; Score 343.5; DB 3; Length 909;
Best Local Similarity 23.2%; Pred. No. 2.8e-23;
Matches 111; Conservative 72; Mismatches 177; Indels 119; Gaps 15;

QY 88 CGHLVYQDSGTMSTKNYPTGTYNHTVCEKTIYPR-GKRLILRLG-DLIDIESQTCASDY 145
DB 28 CGGRPNKSDAGYITSPGYPODYPSHQNCEMIYAPENQKIYLNFPHEIEKHDKYDF 87
QY 146 LLEF-----TSSDQYGPYCGSMTPVKELLNTSEVTVPRESGSHISGRGFLITY-----AS 196
DB 88 IEIRDGDSADLLGKHGNI-APPTIISGSVLYIKFTSDVARQAGFSLEYEIFKTGS 146
QY 197 SD-----HPDLITC-----LERASHYLTKEYSKF----- 220
DB 147 EDCSKNFTSPNGTIESPGPEKYPHNLDCFTFLAKPRMEITLQFLTFDLHDPLOYEGC 206
QY 221 -CPAGRDVAGDIS--GNVVDGYRDTSLCKAIIHAGIIADELGGQISVLRKGISRY-- 275
DB 207 DCKYMDLIDWIGIPHVGLIGKCTKTPSKLRSTGILSTFTTDMAVAKOGFSARYL 266
QY 276 -----EGILANGVLSRDGSLSDKRLFTSNCGCSRLSFEPPDQIRAS 317
DB 267 IHOEPENFOCNVPLGMESGRIANDQISASTFSDGR----- 303
QY 318 SSMQSVNESGQVHMSPGQARLDQGPWSAGDSSNNHPRMELIDLGKKKTIIGIRT 377
DB 304 -----WTPQOSRLHGDDNGWTPMLDSN-----KEYLVDLRLFLMLTAIATQ 345
QY 378 G--STOSNENFYVKSFYVNFKNNSKMTYKGIYNNEEKVFOGNSNFRDPVONNFTPIY 435
DB 346 GAISRETOKGYVKSYLEVSTNGEDMMVYRHGKNH--KIFQANNDAFEVVLNKLHMLPL 403
QY 436 ARYRVVPOTHQRIALVELIGCOITQGNDSLVRKTS---QSTSVSTFKKEDETIRP 491
DB 404 TRFIRIRQTHLIALRLLEFGCRVTDAPCSNMLGLMLGLIADTQISASTREYLWSP 462

RESULT 6
US-08-936-135-10
Sequence 10, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA

COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-10

Query Match 11.3%; Score 343.5; DB 3; Length 909;
Best Local Similarity 23.2%; Pred. No. 2.8e-23;
Matches 111; Conservative 72; Mismatches 177; Indels 119; Gaps 15;

QY 88 CGHLVYQDSGTMSTKNYPTGTYNHTVCEKTIYPR-GKRLILRLG-DLIDIESQTCASDY 145
DB 28 CGGRPNKSDAGYITSPGYPODYPSHQNCEMIYAPENQKIYLNFPHEIEKHDKYDF 87
QY 146 LLEF-----TSSDQYGPYCGSMTPVKELLNTSEVTVPRESGSHISGRGFLITY-----AS 196
DB 88 IEIRDGDSADLLGKHGNI-APPTIISGSVLYIKFTSDVARQAGFSLEYEIFKTGS 146
QY 197 SD-----HPDLITC-----LERASHYLTKEYSKF----- 220
DB 147 EDCSKNFTSPNGTIESPGPEKYPHNLDCFTFLAKPRMEITLQFLTFDLHDPLOYEGC 206
QY 221 -CPAGRDVAGDIS--GNVVDGYRDTSLCKAIIHAGIIADELGGQISVLRKGISRY-- 275
DB 207 DCKYMDLIDWIGIPHVGLIGKCTKTPSKLRSTGILSTFTTDMAVAKOGFSARYL 266
QY 276 -----EGILANGVLSRDGSLSDKRLFTSNCGCSRLSFEPPDQIRAS 317
DB 267 IHOEPENFOCNVPLGMESGRIANDQISASTFSDGR----- 303
QY 318 SSMQSVNESGQVHMSPGQARLDQGPWSAGDSSNNHPRMELIDLGKKKTIIGIRT 377
DB 304 -----WTPQOSRLHGDDNGWTPMLDSN-----KEYLVDLRLFLMLTAIATQ 345
QY 378 G--STOSNENFYVKSFYVNFKNNSKMTYKGIYNNEEKVFOGNSNFRDPVONNFTPIY 435
DB 346 GAISRETOKGYVKSYLEVSTNGEDMMVYRHGKNH--KIFQANNDAFEVVLNKLHMLPL 403
QY 436 ARYRVVPOTHQRIALVELIGCOITQGNDSLVRKTS---QSTSVSTFKKEDETIRP 491
DB 404 TRFIRIRQTHLIALRLLEFGCRVTDAPCSNMLGLMLGLIADTQISASTREYLWSP 462

RESULT 7
US-08-936-135-12
Sequence 12, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang

TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 914 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
OS-08-936-135-12

Qy 436 ARYRVVQTMORALAVELICQITQGNDSLWKRKTS---QSTSVSTKKEDETTIRP 491
Db 404 TRFIRIRPQTWHLGIALRLLELFGCRVTDAPCSNMLGMLSGLIADVOISASSTREYLMS 462

RESULT 9

US-08-936-135-16
Sequence 16, Application US/08936135
Patent No. 6054293

GENERAL INFORMATION:

APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 931 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-16

Query Match 11.3%; Score 343.5; DB 3; Length 931;
Best Local Similarity 23.2%; Pred. No. 3e-23;
Matches 111; Conservative 72; Mismatches 177; Indels 119; Gaps 15;

Qy 88 CGHLVTVQDSGTMSTKNKPGTYPRNTHVEKITIYPK-GKRLILRLG-DLDIESOTCASDY 145
Db 28 CGGRPNSSDAGITSPGYPDYPHONCEWTVYAPENQKVLNPNPHEIEKHKCKYDF 87
Qy 146 LLF-----TSSSDOYGYPCGSMTPKELLNLTSEVTVREFSGSHISGRFLLTY-----AS 196
Db 88 IEIRGDESADLGRKHCNLT-APPTLISSGSVLTKFTSDYANAGAFSLRYELFKGS 146
Qy 197 SD-----HPDLITC-----LERASHYKTEYSKF----- 220
Db 147 EDCSKNFTSPNGTIESPGPEKYPHNLDCTFTLAKPRMELLQFLTFDEHDPLOYEG 206
Qy 221 -CPAGCRVYADDIS--GMMVGYRDTSLILCAAHAGIIMBELGQISVLDORKGISRI-- 275
Db 207 DCKTDWLDIMDGIPIHVGILKCYGCTKTPSKLRSSSTGLSLFTTDMVAVADGFSARYYL 266
Qy 276 -----EGILANGVLSRDGSLSDKRFLLFTSNCGSRSLSPDPGQIRAS 317
Db 267 IHQEPENFGCNVPLGMESGRILANDQISASSTPSDGR----- 303

Qy 318 SSMQSVNESGDVHMSPCQARLDQGPSMASGSSNNHKPREMLEIDLEKKKITIGIRKT 377
Db 304 -----WTPOOSRLHGDDNCGWTPNLDN-----KEYLVQDILRFLTMLTALATQ 345
Qy 378 G--STQSNFNFYVKSFVNFKNNSKMTYKGIYVNEEKVQGSNFRDPVQNNFPIPIV 435
Db 346 GAISRETQKGYVYKSLKEVSTNGEDMMVYRHKNH--KIFQANNDADEVYLNKLHMDLL 403
Qy 436 ARYRVVQTMORALAVELICQITQGNDSLWKRKTS---QSTSVSTKKEDETTIRP 491
Db 404 TRFIRIRPQTWHLGIALRLLELFGCRVTDAPCSNMLGMLSGLIADVOISASSTREYLMS 462

RESULT 10

US-08-936-135-6
Sequence 6, Application US/08936135
Patent No. 6054293

GENERAL INFORMATION:

APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 923 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-6

Query Match 10.8%; Score 329; DB 3; Length 923;
Best Local Similarity 24.0%; Pred. No. 6.8e-22;
Matches 113; Conservative 79; Mismatches 161; Indels 118; Gaps 21;

Qy 64 RGL-----LALLAVSAPRLQAEELGDCGHLVTVQDSGTMSTKNKPGTYPRNTHVCEK 117
Db 3 RGLPLLCATLALMALAGAFR-----SDKCGGTIKLENPGVLYLSPGYSHYPSKCEW 56
Qy 118 TIYVPGK-KRLILRLG-DLDIESOTCASDYLLTSSSDO-----YGPYCGSMTPKELLN 171
Db 57 LIOAPEYQRIIINFNPFLIEDRDCKYDYVEVYIDGENEGRLMKFGCKRI-APSPVSS 115
Qy 172 TSEVTVREFSGSHISGRFLLTYA-----SSDHPDLITC-- 205
Db 116 GPFLLFKFVSDYETHGAGSIRIETIRKRGPCSONTAPRGVYKSPGPEKYPNCLECY 175
Qy 206 -----LERASHYKTEYSKFCPCAG--CRDVAAGDISGMNVGVRTDLSILCAAH 252

Db 176 IIRAPKSEITLLEFESFLEKDSNP--PGMFRTDRLLE---WDGFPE-----VGRH 223
QY 253 AG-11ADELGGOI-----SVLQRKGISRYEILANGVLSRDSLSDRKFLF 297
Db 224 IGRGCGKTPGRIRSSGVLWVFYTDASIAKGEFSANYSVLQSSI-SEDFK-----274
QY 298 TSNCGSLSFEF---DGIIRASSSMQSVNEGSDVHWPQARLDQGPMSAGSDSN 353
Db 275 ---CMALCMESGEIHSDDITASSOYGT-----NMSVERSLNYPENGWTPGEDSY 322
QY 354 NHRPREMLEIDLEKKKITGIRTTG--STQSNFNFYKSFVNMFNKNNKSKTKYGIYNN 411
Db 323 ---KENIQVDLGLIRVTVAGTQGAISKETKKYIKYTRVIDISSNGEEMISLK--EGN 376
QY 412 EEKVFQGSNFRDPVQNNFIPPIVARYRVVPTQHORIALKVELIGCOIT 462
Db 377 KAIIFQGNTPDVLGLVFSKPLITRFVRIRKPYMSMETGISMRREVYCKKIT 427
RESULT 11
US-07-607-538C-4
; Sequence 4, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; APPLICANT: Laroocca, David J.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Amzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA
; ZIP: 94596
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/607,538C
; FILING DATE: 01-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Viviana Amzel
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: CRFCC-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1931
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
US-07-607-538C-4
Query Match 10.5%; Score 321; DB 1; Length 218;
Best Local Similarity 42.6%; Pred. No. 3.5e-22;

Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;
QY 300 NGCSRSLSFE---PDGQIRASS---SMQSVNEGSDVHWPQARLDQGPMSAGSDS 352
Db 58 NGCSSTPLGMENKRIENKQITASSFKKSMW---GD--YWEPPRALNAGRVNMAQAKA 110
QY 353 NHRPREMLEIDLEKKKITGIRTTG--STQSNFNFYKSFVNMFNKNNKSKTKYGIYNN 412
Db 111 NNKK--QWLEIDLTKTKITAITITGCKSLSEDMYKSVTHYSDGVEMKPYRLKSSMV 168
QY 413 EEKVFQGSNFRDPVQNNFIPPIVARYRVVPTQHORIALKVELIGCOI 461
Db 169 DKIFEGNTWTKGHVKNFNPPIISRPIRVIPRTWNSIALRIELRGCDI 217
RESULT 12
US-08-162-402B-4
; Sequence 4, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-4
Query Match 10.5%; Score 321; DB 2; Length 218;
Best Local Similarity 42.6%; Pred. No. 3.5e-22;
Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;
QY 300 NGCSRSLSFE---PDGQIRASS---SMQSVNEGSDVHWPQARLDQGPMSAGSDS 352
Db 58 NGCSSTPLGMENKRIENKQITASSFKKSMW---GD--YWEPPRALNAGRVNMAQAKA 110
QY 353 NHRPREMLEIDLEKKKITGIRTTG--STQSNFNFYKSFVNMFNKNNKSKTKYGIYNN 412

Db 111 NNKK--QWLEIDLKIKITAITITQCKSLSEMYKSYTIHYSEQVEKMPYRLKSSMV 168
QY 413 EKVFGNSNFRDPYQNNFIPPIYARVRYVPQVTHQRIALVELIGCQI 461
Db 169 DLFEGSNNTKGHWKNEFPPIISRIRIPIKTNOSIALRLLEFGCDI 2117

RESULT 13

US-08-746-111-5
Sequence 5, Application US/08746111
Patent No. 6066778
GENERAL INFORMATION:
APPLICANT: Ginsburg, David
APPLICANT: Cui, Jisong
TITLE OF INVENTION: Compositions And Methods For Screening
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,111
FILING DATE: 06-NOV-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2183 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-746-111-5

Query Match 10.4%; Score 318; DB 3; Length 2183;
Best Local Similarity 40.8%; Pred. No. 3,1e-20;
Matches 69; Conservative 31; Mismatches 53; Indels 16; Gaps 5;
QY 300 NCGSRSLSF-----PDQIRASS---SMQSYNESGDVHNSPGQARIQDQGPSWASGSDSS 352
Db 2023 NCGSTPLGLEGRIGDQIRASSFKKSW-----GD--YWEPPARLNAQGRVNAQAKAN 2075
QY 353 NHRPREMLEIDGEEKKKTIGRTTGSTQSNFNFYKSFVWNEKNNNSKKTYKGIYNNE 412
Db 2076 NNNK--QWLDYDLKIKITAIITQCKSLSEMYKSYTIHYSDQGVAMKPYRLKSSMV 2133
QY 413 EKVFGNSNFRDPYQNNFIPPIYARVRYVPQVTHQRIALVELIGCQI 461
Db 2134 DKIEGNSNNTKGHWKNEFPPIISRIRIPIKTNOSIALRLLEFGCDI 2182

RESULT 14
US-08-162-402B-14
Sequence 14, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.

APPLICANT: IAROCCA, DAVID J.
TITLE OF INVENTION: 46 KOALTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 160 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-14

Query Match 10.2%; Score 311; DB 2; Length 160;
Best Local Similarity 41.7%; Pred. No. 1,8e-21;
Matches 70; Conservative 30; Mismatches 52; Indels 16; Gaps 5;
QY 301 GCSRSLSFE-----PDQIRASS---SMQSYNESGDVHNSPGQARIQDQGPSWASGSDSSN 353
Db 1 GCSSTPLGEMENKIEKQIRASSFKKSW-----GD--YWEPPARLNAQGRVNAQAKAN 53
QY 354 NHRPREMLEIDGEEKKKTIGRTTGSTQSNFNFYKSFVWNEKNNNSKKTYKGIYNNEE 413
Db 54 NNNK--QWLEIDLKIKITAIITQCKSLSEMYKSYTIHYSDQGVAMKPYRLKSSMVD 111
QY 414 KTVQGNSNFRDPYQNNFIPPIYARVRYVPQVTHQRIALVELIGCQI 461
Db 112 KIEGNTNKGHWKNEFPPIISRIRIPIKTNOSITLRLLEFGCDI 159

RESULT 15
US-09-324-867-2
Sequence 2, Application US/09324867A
Patent No. 6251632
GENERAL INFORMATION:
APPLICANT: Lallierap, David
APPLICANT: Cameron, Cherie
APPLICANT: No. 6251632ley, Colleen
APPLICANT: Horrocks, L. Suzanne Hoyle
APPLICANT: Hough, Christine
TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
FILE REFERENCE: 1669,0010002/JAG/BJD
CURRENT APPLICATION NUMBER: US/09/324,867A
CURRENT FILING DATE: 1999-06-03
EARLIER APPLICATION NUMBER: 09/035,141

; EARLIER FILING DATE: 1998-03-059
; EARLIER APPLICATION NUMBER: 60/039,953
; EARLIER FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2343
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-324-867-2

Query Match 10.0%; Score 306.5; DB 4; Length 2343;
Best Local Similarity 40.5%; Pred. No. 4.2e-19;
Matches 68; Conservative 26; Mismatches 61; Indels 13; Gaps 4;

QY 300 NCGRSLSEFE---PDGQIRASSSMQSVNESGDOVHMSPGOARLQDOGSPSWASGDSSNNH 355
Db 2183 NCSCMPLGMEKRAISDAQITASSYLSSMLAT----WSPSOARLHLCRTNWMRPQANN- 2236
QY 356 KPREMLEIDLGEKKKITGIRPTGSGTQSNFNFEYKSFVNMFKNNSKWKTYKGIYNNNEKV 415
Db 2237 -PKRELQYDFEKKTKYKVCITTOGVKSLISMVKEFLISSQDGHNMTLF--LQNGKYKV 2293
QY 416 FQGSNFRDPVQNNFIPPIVARYRVVYQVTHORIALKVELIGCOITQ 463
Db 2294 FQGNRDSSTPYNRLEPLVARYVRLHPOSMAHIALRLAEVLGCDTQO 2341

Search completed: July 9, 2002, 02:15:21
Job time: 3788 sec

Db 1621 gaggctcaaaaacagactgttgaagcagatataatcccttgcagacatcagta 1680
Qy 1681 gctgagttaccatcagctatagataagagagatgacacaaattgattcctc 1740
Db 1681 gctgagttaccatcagctatagataagagagatgacacaaattgattcctc 1740
Qy 1741 acaagtgatattgacagattaa 1761
Db 1741 acaagtgatattgacagattaa 1761

RESULT 2

AA000660 standard; cDNA: 3594 BP.

AA000660;

07-SEP-2001 (first entry)

Human TANGO 229 cDNA.

Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung; skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte; bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer; anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss; malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis; attention deficit disorder; Crohn's disease; gastroenteritis; goitre; hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism; muscular dystrophy; immuno-competence; vertebrate; blood; serum.

Homo sapiens.

Location/Qualifiers

72..2219

/*tag- a

/product= "Human TANGO 229"

72..173

/*tag- b

174..2216

/*tag- c

/product= "Mature human TANGO 229"

WO000129088-A1.

26-APR-2001.

23-JUN-2000; 2000MO-US17386.

19-OCN-1999; 99US-0420707.

(MILL-) MILLENNIUM PHARM INC.

Mockay CR, Myers PS, KIRST SJ, Fraser CC, Leiby KR;

WPI: 2001-308477/32.

P-PSDB; AAU00670.

New isolated nucleic acid molecule for diagnosis, prevention, and therapy of human and other animal disorder, or as modulating agent for regulating cellular processes

Claim 1; Fig 1; 263pp; English.

This sequence represents a cDNA which encodes human TANGO 229 polypeptide. The protein and similar others exhibit the ability to affect growth, proliferation, survival, differentiation, activity, morphology, or movement/migration of, e.g., T cells and cells of the heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus tissue. They can be used as modulating agents for regulating cellular processes, thus, the proteins and their associated nucleic acids can be used to prognosticate, prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma,

CC anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis, CC meningitis, attention deficit disorder, Crohn's disease, gastroenteritis, CC goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and muscular dystrophy. Antibodies to disorders such as these CC can be made by providing a polypeptide of the invention to an CC immuno-competent vertebrate and harvesting blood or serum from the CC vertebrate.

Sequence 3594 BP; 958 A; 810 C; 921 G; 905 T; 0 other;

Query Match 95.8%; Score 1686.8; DB 22; Length 3594;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1688; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 71 ggcggggcagctgagctcgagatccgtcgagggagagcagcttgcagactgagc 130
Db 1 ggcggggcagctgagctcgagatccgtcgagggagagcagcttgcagactgagc 60
Qy 131 ccagcgagctatgctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 190
Db 61 ccagcgagctatgctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
Qy 191 gggggcccgagcttgcctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 250
Db 121 gggggcccgagcttgcctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180
Qy 251 tgggtgagctgtgagcactgagctatcagagatgagatgagatgagatgagatgagat 310
Db 181 tgggtgagctgtgagcactgagctatcagagatgagatgagatgagatgagatgagat 240
Qy 311 attatccggagactaccccaatcacactgttgcgaaagaacattacagatccaaag 370
Db 241 attatccggagactaccccaatcacactgttgcgaaagaacattacagatccaaag 300
Qy 371 ggaagaacatgattcaggttggagatttgatagatccagacactgtgcttctg 430
Db 301 ggaagaacatgattcaggttggagatttgatagatccagacactgtgcttctg 360
Qy 431 actatctcttcacacagctcttcacacacacacacacacacacacacacacacacacac 490
Db 361 actatctcttcacacagctcttcacacacacacacacacacacacacacacacacacac 420
Qy 491 ttcccaagaacatcgttgaacaagaagtaaacgcgttcgcttgaagtgatccc 550
Db 421 ttcccaagaacatcgttgaacaagaagtaaacgcgttcgcttgaagtgatccc 480
Qy 551 acattctgcgcgggttttctgctgacatcgagcagcagcagcagcagcagcagcagcagc 610
Db 481 acattctgcgcgggttttctgctgacatcgagcagcagcagcagcagcagcagcagcagc 540
Qy 611 catgttggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 670
Db 541 catgttggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
Qy 671 gtttgagagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 730
Db 601 gtttgagagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
Qy 731 cttatttgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 790
Db 661 cttatttgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
Qy 791 tcaagtgtcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 850
Db 721 tcaagtgtcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 780
Qy 851 ttccggagagctgtcccgtagaagaacagcagcagcagcagcagcagcagcagcagcagcagc 910
Db 781 ttccggagagctgtcccgtagaagaacagcagcagcagcagcagcagcagcagcagcagcagc 840
Qy 911 gatccttgagcttgaacatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 970

Db	841	gaccccttgagatcttgcgaacctgcagcgcaaaatcagagcctctcttcattggtcagtcgtca	900
Qy	971	atgagagctggagagcccaagcttcacTgctgcctcTgcgcgaagcccgaccttcagagcccaagcc	1030
Db	901	atgagagctggagagcccaagcttcacTgctgcctcTgcgcgaagcccgaccttcagagcccaagcc	960
Qy	1031	calcatagggcttcTgcgcgcagctatgacacaaacccaagagctggctgtgagatctg	1090
Db	961	calcatagggcttcTgcgcgcagctatgacacaaacccaagagctggctgtgagatctg	1020
Qy	1091	attctggggagaaaagaaaataaacagagatattagagccacagagatctctacacagtcgact	1150
Db	1021	attctggggagaaaagaaaataaacagagatattagagccacagagatctctacacagtcgact	1080
Qy	1151	tcaactcttatgtgttaagagttttgtgagtgaacttcaaaaaaataattcttagtgagaga	1210
Db	1081	tcaactcttatgtgttaagagttttgtgagtgaacttcaaaaaaataattcttagtgagaga	1140
Qy	1211	cctataaagagaaattctgcgaataatgaaagagtgcttcagagtgaaactctaaactcttcggg	1270
Db	1141	cctataaagagaaattctgcgaataatgaaagagtgcttcagagtgaaactctaaactcttcggg	1200
Qy	1271	accacagtcgcaaaaacattctcaTccctcccatctgctgcgcagatattgctgggtgttccccc	1330
Db	1201	accacagtcgcaaaaacattctcaTccctcccatctgctgcgcagatattgctgggtgttccccc	1260
Qy	1331	agacatgcccacagagatagcctctgaaagtgagatctcatctggtctgcacagatctaacagaag	1390
Db	1261	agacatgcccacagagatagcctctgaaagtgagatctcatctggtctgcacagatctaacagaag	1320
Qy	1391	gtaatgattctattgggtggtgcgcgaagacaagctcaagacaacagtgcttccaactaagaag	1450
Db	1321	gtaatgattctattgggtggtgcgcgaagacaagctcaagacaacagtgcttccaactaagaag	1380
Qy	1451	aagaatgagacaatctcacagaagcccatccctctgcgaaagaaatccacagagaaataaacatctta	1510
Db	1381	aagaatgagacaatctcacagaagcccatccctctgcgaaagaaatccacagagaaataaacatctta	1440
Qy	1511	caaaagctggtattccattcttggtgcctcttggtgcctcgtgtgttgcgtgaattgggattc	1570
Db	1441	caaaagctggtattccattcttggtgcctcttggtgcctcgtgtgttgcgtgaattgggattc	1500
Qy	1571	ctgcagagcctcttagaagaagaagaagaagaagagctccgtatctgatactgacagcyagagctcaga	1630
Db	1501	ctgcagagcctcttagaagaagaagaagaagaagagctccgtatctgatactgacagcyagagctcaga	1560
Qy	1631	aaacagagcctgttggaagcagatcttaaatatccctcttgccagacatacagctcagctgagttca	1690
Db	1561	aaacagagcctgttggaagcagatcttaaatatccctcttgccagacatacagctcagctgagttca	1620
Qy	1691	ccatcacagctatgaataagagaagagaatgacacaaaagttagatctcatcacaaagtgtata	1750
Db	1621	ccatcacagctatgaataagagaagagaatgacacaaaagttagatctcatcacaaagtgtata	1680
Qy	1751	tggcagagttta 1760	
Db	1681	tggcagagttta 1690	

RESULT 3
AAS00615
ID AAS00615 standard; cDNA; 1620 BP.
XX
XX
AC AAS00615;
XX
DT 29-AUG-2001 (first entry)
XX
XX
DE Novel human protein (NHP) DNA sequence #3.
XX
KM Novel human protein; NHP; CUB domain; extracellular domain; gene therapy.
KM obesity; high blood pressure; connective tissue disorder; infertility;
KM NHP-mediated pathway; ss.

XX	Homo sapiens.
OS	
XX	
XX	Key Location/Qualifiers
FE	1..1620
FE	CDS
FE	/tag= A
FE	/product= "Novel human protein #3"
XX	
XX	WO200129219-A1.
PN	
PD	26-APR-2001.
XX	
XX	08-OCT-2000; 2000WO-US28798.
PF	
XX	
PR	19-OCT-1999; 99US-0160285.
PR	18-FEB-2000; 2000US-0183583.
PA	(LEXI-) LEXICON GENETICS INC.
PI	
DR	Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX	
XX	WPI: 2001-290917/30.
DR	P-PSDB; AAU00630.
PT	
PT	Novel nucleic acid encoding human CUB-domain containing protein, useful
PT	for drug screening, diagnosis and treatment of physiological disorders
XX	or diseases -
PS	
XX	Disclosure: Page 28-29; 33pp; English.
CC	
CC	The sequence represents a polynucleotide which encodes a novel human
CC	protein (NHP) containing a CUB domain (an extracellular domain). CUB
CC	proteins have been associated with regulating development, modulating
CC	cellular processes and preventing infectious disease. NHP nucleotide
CC	sequences are useful for gene therapy of physiological disorders or
CC	diseases. NHP oligonucleotides are useful as hybridisation probes for
CC	screening libraries and assessing gene patterns. NHP nucleotide sequences
CC	are useful for detecting mutant or inappropriately expressed NHPs (for
CC	example, those proteins associated with obesity, high blood pressure,
CC	connective tissue disorders and infertility) for the diagnosis of a
CC	disease. The polynucleotides may also be used in screening for drugs
CC	effective in the treatment of symptomatic or phenotypic manifestations of
CC	perturbing the normal function of NHP in the body. Nucleotide constructs
CC	encoding NHP products are used to genetically engineer host cells to
CC	express such products in vivo. These host cells allow for the
CC	identification of compounds that bind to NHP receptors or trigger
CC	NHP-mediated pathways.
SO	
XX	
XX	Sequence 1620 BP; 456 A; 353 C; 422 G; 389 T; 0 other;
Query Match	92.0%; Score 1620; DB 22; Length 1620;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1620; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	142 atggtgccgcggcgccgcggcggcgacgtcgcggtcgctcgcgcgccctctctg 201
Db	1 atggtgccgcggcgccgcggcggcgacgtcgcggtcgctcgcgcgccctctctg 60
OY	202 gtttgctgctcgcggtcttcgcgcgcgcgtcgctcgcaaggaggaagtgtgatgac 261
Db	61 gcttgcgtcgcggtcttcgcgcgcgcgtcgctcgcaaggaggaagtgtgatgac 120
OY	262 ttggagcacactgtagctattatcagaatatgtgacaataaacatctaagaatatcccgag 321
Db	121 ttgtgagcacactgtagctattatcagaatatgtgacaataaacatctaagaatatcccgag 180
OY	322 accaccaccaatacacacctgtttgcgaaagacaattaacgttacccaaggggaaaagactg 381
Db	181 acctaccaccaatacacacctgtttgcgaaagacaattaacgttacccaaggggaaaagactg 240
OY	382 attctgaggttggagatttgtgaatcatccagacctgtgctctgaactctctccc 441

Db 241 attctgaggttgaggatcttgatatacgaatccagacctgtgctctctactatctctc 300
 QY 442 ttcacacacctcttcagatcaaatatggtccatcactgttgaaatagactggtcccaagaa 501
 Db 301 ttcaacagctcttcagatcaaatatggtccatcactgttgaaatagactggtcccaagaa 360
 QY 502 ctctgttgaaacacaagtgaagtaaccgtccgtcttgagagtgagatccacattctgac 561
 Db 361 ctctgttgaaacacaagtgaagtaaccgtccgtcttgagagtgagatccacattctgac 420
 QY 562 cggggttttttgcgtactatgacagacagacacacacacacacacacacacacacacac 621
 Db 421 cggggttttttgcgtactatgacagacagacacacacacacacacacacacacacacac 480
 QY 622 cgaactagccattatgaaagacagaaatagacagaaatcttcgacgtctgtgtgagagac 681
 Db 481 cgaactagccattatgaaagacagaaatagacagaaatcttcgacgtctgtgtgagagac 540
 QY 682 gtagcagaagacacattctctggaatatagtagatagataagaaatcctcttcttctgc 741
 Db 541 gtagcagaagacacattctctggaatatagtagatagataagaaatcctcttcttctgc 600
 QY 742 aaagctgcacatccatgacagaaatattgctgataagtaagtgagcagatcagtgctt 801
 Db 601 aaagctgcacatccatgacagaaatattgctgataagtaagtgagcagatcagtgctt 660
 QY 802 cagcgcaaaaggagatcagatcagatagagagatctgcgcacatggtgtctcttcgaaggat 861
 Db 661 cagcgcaaaaggagatcagatcagatagagagatctgcgcacatggtgtctcttcgaaggat 720
 QY 862 ggttcctctgcaagaacagcagattctgttacctccaaatggtgagcagatccttgagt 921
 Db 721 ggttcctctgcaagaacagcagattctgttacctccaaatggtgagcagatccttgagt 780
 QY 922 ttgaaacctgacgggcaaatcagaagctcttctccatgacagatcagtgagatgagatgga 981
 Db 781 ttgaaacctgacgggcaaatcagaagctcttctccatgacagatcagtgagatgagatgga 840
 QY 982 gaccaagtctactggtctcctgccaagccgacattcagagcaaacagcccatatgggt 1041
 Db 841 gaccaagtctactggtctcctgccaagccgacattcagagcaaacagcccatatgggt 900
 QY 1042 tcgggagacagtgacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 1101
 Db 901 tcgggagacagtgacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 960
 QY 1102 aaaaagaataacaggaattagacacagatctacacagcagaaactcaacttcat 1161
 Db 961 aaaaagaataacaggaattagacacagatctacacagcagaaactcaacttcat 1020
 QY 1162 gttaaagatttgtagtaactcaaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 1221
 Db 1021 gttaaagatttgtagtaactcaaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 1080
 QY 1222 attgtaataatgaagaaaggtgttcagggtaacttaactttcgggacccagtgcaa 1281
 Db 1081 attgtaataatgaagaaaggtgttcagggtaacttaactttcgggacccagtgcaa 1140
 QY 1282 aacaatttcatccctccatcgtgacagataatgtagcggtgtgtccccaagacatgacac 1341
 Db 1141 aacaatttcatccctccatcgtgacagataatgtagcggtgtgtccccaagacatgacac 1200
 QY 1342 cagaagatagccttgaaagtgagatcattggttgcagaaatatacagaagtaatgata 1401
 Db 1201 cagaagatagccttgaaagtgagatcattggttgcagaaatatacagaagtaatgata 1260
 QY 1402 ttggttgacgacagacagac 1461
 Db 1261 ttggttgacgacagacagac 1320
 QY 1462 atcaacaaggccatccctcggagaagaacatcacacagaaataaacattcaacacggtgct 1521
 Db 1321 atcaacaaggccatccctcggagaagaacatcacacagaaataaacattcaacacggtgct 1380

QY 1522 attccattggtgctcctctgtctgtcctgctggtgttgcgtggaatgggagatcttcagacctt 1581
 Db 1381 attccattggtgctcctctgtctgtcctgctggtgttgcgtggaatgggagatcttcagacctt 1440
 QY 1582 agaaagaagaagaagaagaagatccgtatgatacagcgaagctcagaagaacagactgt 1641
 Db 1441 agaaagaagaagaagaagaagatccgtatgatacagcgaagctcagaagaacagactgt 1500
 QY 1642 tggaaagcagataataatcccttgcagacatcagtcagctgagtttaaccatcagat 1701
 Db 1501 tggaaagcagataataatcccttgcagacatcagtcagctgagtttaaccatcagat 1560
 QY 1702 gataatgagaagagatgatacaaaaagttgagatcattcaacaagtatgtagcagttta 1761
 Db 1561 gataatgagaagagatgatacaaaaagttgagatcattcaacaagtatgtagcagttta 1620

RESULT 4
 AAS00613
 ID AAS00613 standard; DNA: 1768 BP.
 AC AAS00613;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Novel human protein (NHP) DNA sequence #1.
 XX
 KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
 KW obesity; high blood pressure; connective tissue disorder; infertility;
 KW NHP-mediated pathway; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 59..1522
 FT /tag= a
 FT /product= "Novel human protein #1"
 XX
 PN WO200129219-A1.
 XX
 PD 26-APR-2001.
 XX
 PF 08-OCT-2000; 2000WO-US28798.
 XX
 PR 19-OCT-1999; 99US-0160285.
 PR 18-FEB-2000; 2000US-0183583.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 DR WPI. 2001-290917/30.
 DR P-PSDB; AAU00628.
 XX
 PT Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 XX
 PS Disclosure: Page 30-31; 33pp; English.
 XX
 CC The sequence represents a polynucleotide which encodes a novel human
 CC protein (NHP) containing a CUB domain (an extracellular domain). CUB
 CC proteins have been associated with regulating development, modulating
 CC cellular processes and preventing infectious disease. NHP nucleotide
 CC sequences are useful for gene therapy of physiological disorders or
 CC diseases. NHP oligonucleotides are useful as hybridisation probes for
 CC screening libraries and assessing gene patterns. NHP nucleotide sequences
 CC are useful for detecting mutant or inappropriately expressed NHPs (for
 CC example, those proteins associated with obesity, high blood pressure,
 CC connective tissue disorders and infertility) for the diagnosis of a
 CC disease. The polynucleotides may also be used in screening for drugs
 CC effective in the treatment of symptomatic or phenotypic manifestations of

SEQ ID NO:3
Page 6
AC NO: AAA96736, Database: N-gene, -032802

CC perturbing the normal function of NHP in the body. Nucleotide constructs
CC encoding NHP products are used to genetically engineer host cells to
CC express such products in vivo. These host cells allow for the
CC identification of compounds that bind to NHP receptors or trigger
CC NHP-mediated pathways.

SQ Sequence 1768 BP; 539 A; 370 C; 421 G; 438 T; 0 other;

Query Match	86.4%	Score 1523	DB 22	Length 1768
Best Local Similarity	100.0%	Pred. No.	0	
Matches 1523	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

QY	240	gggagagagactgggtgagabgcgtgtgagaaactaaatgacttaacagagatagtgacaaat	299
Db	1	ggcgagagagacgcgggtgagabgcgtgtgagaaactaaatgacttaacagagatagtgacaaat	60
QY	300	gacacttaagaatathatcccgagacactaccaccaatacacgtgttgagaaaagacaattac	359
Db	61	gacacttaagaatathatcccgagacactaccaccaatacacgtgttgagaaaagacaattac	120
QY	360	agttccaaagggggaagaagacttaattcttgaggtttggagagattggatccgaatccagac	419
Db	121	agttccaaagggggaagaagacttaattcttgaggtttggagagattggatccgaatccagac	180
QY	420	ctgtgcttcctgacatattctctctcaacccagctcttcagaatcaatattggtccatacttgg	479
Db	181	ctgtgcttcctgacatattctctctcaacccagctcttcagaatcaatattggtccatacttgg	240
QY	480	aagtatgactggttcccgaagaataactctgtttgaacacaaagtgaagtaaacggtccgcttga	539
Db	241	aagtatgactggttcccgaagaataactctgtttgaacacaaagtgaagtaaacggtccgcttga	300
QY	540	gagtgagatccacaactctctcgccggggttttttgcgcgaactctggagagacgacacatcc	599
Db	301	gagtgagatccacaactctctcgccggggttttttgcgcgaactctggagagacgacacatcc	360
QY	600	agatttaataacatgctttgttgaaacgagctagccattatcttgaaagacagaatacagcaaat	659
Db	361	agatttaataacatgctttgttgaaacgagctagccattatcttgaaagacagaatacagcaaat	420
QY	660	ctgcccacactgtgtgtgagagacgttagcaagagacattctcttgagaaatagtgatagtgata	719
Db	421	ctgcccacactgtgtgtgagagacgttagcaagagacattctcttgagaaatagtgatagtgata	480
QY	720	tagagataactctcttattgtgtgcaaaagctgcacatcagatcagaataattgtctgaatgaact	779
Db	481	tagagataactctcttattgtgtgcaaaagctgcacatcagatcagaataattgtctgaatgaact	540
QY	780	aggttgagccaagatcaagtgctgtctcaacgcgcaaaagagatcagatcagatcttgaaaggatcttgc	839
Db	541	aggttgagccaagatcaagtgctgtctcaacgcgcaaaagagatcagatcagatcagatcttgc	600
QY	840	caatgctgtctcttcgagagagatcagttccctctgacagaacaagcgagatctctgttttaccctcaa	899
Db	601	caatgctgtctcttcgagagagatcagttccctctgacagaacaagcgagatctctgttttaccctcaa	660
QY	900	tgtgttgcaagcagatcccttgagatttttgaacctgcgggcgaatcaagagcttcttccatactg	959
Db	661	tgtgttgcaagcagatcccttgagatttttgaacctgcgggcgaatcaagagcttcttccatactg	720
QY	960	gagatcgagatcaatagaggttgagacccaagtctcaactgtctcctcgagccaagcccgacttca	1019
Db	721	gagatcgagatcaatagaggttgagacccaagtctcaactgtctcctcgagccaagcccgacttca	780
QY	1020	ggagccaagggccatcatatgggcttctcgggcagctagagaaacaaaccaacaaaccaagagtg	1079
Db	781	ggagccaagggccatcatatgggcttctcgggcagctagagaaacaaaccaacaaaccaagagtg	840
QY	1080	gctgtgagatcgaatttggggggggaagaaagaataaacaaggaattagagacacagaggtctac	1139
Db	841	gctgtgagatcgaatttggggggggaagaaagaataaacaaggaattagagacacagaggtctac	900

Oy	1140	acagcgcgaacttcaactttatgtttaagaagttctgtgtgaacttcaaaaataaattc	1199
Db	901	acagtcgaacttcaactttatgtttaagaagttctgtgtgaacttcaaaaataaattc	960
Oy	1200	taagtcgaagacctataaaggaaattgtgtataatgtagaagaaagtgcttcagggttaac	1259
Db	961	taagtcgaagacctataaaggaaattgtgtataatgtagaagaaagtgcttcagggttaac	1020
Oy	1260	taactttgggagcccaagtcgcaaaaataatttcaatccctcccatctgctgcgcagataatgtcg	1319
Db	1021	taactttgggagcccaagtcgcaaaaataatttcaatccctcccatctgctgcgcagataatgtcg	1080
Oy	1320	ggtgtcccccagacatgtgcaccagagataagccttgaaagtgtagagctcatgtgtgcga	1379
Db	1081	ggtgtcccccagacatgtgcaccagagataagccttgaaagtgtagagctcatgtgtgcga	1140
Oy	1380	gattacacaaagtgatattcatgtgtgtgtgtgcgaagaagtcanaagcacagaftgttc	1439
Db	1141	gattacacaaagtgatattcatgtgtgtgtgtgcgaagaagtcanaagcacagaftgttc	1200
Oy	1440	aactagaagaagatgtagacacatacacaagcccaatccctctggaagaacaatccacagg	1499
Db	1201	aactagaagaagatgtagacacatacacaagcccaatccctctggaagaacaatccacagg	1260
Oy	1500	aataaacattacaacggctgtgcattccatctgtgtctcttgtgtccctgtgtgtgtgtgcagg	1559
Db	1261	aataaacattacaacggctgtgcattccatctgtgtctcttgtgtccctgtgtgtgtgtgcagg	1320
Oy	1560	aattgggaatcttgcagcctctttagaagaagaagaagaagaagaagtcgcgtacgtacagc	1619
Db	1321	aattgggaatcttgcagcctctttagaagaagaagaagaagaagaagtcgcgtacgtacagc	1380
Oy	1620	ggaagctcagaagaacagactgtttggaagcagattaaataatcccttgcagacatacagtc	1679
Db	1381	ggaagctcagaagaacagactgtttggaagcagattaaataatcccttgcagacatacagtc	1440
Oy	1680	agctgaagttacatcagatcgtatgatataagaagaagatgcacacaaagttagatctcat	1739
Db	1441	agctgaagttacatcagatcgtatgatataagaagaagatgcacacaaagttagatctcat	1500
Oy	1740	cacaagtgatattgcaggttaa	1761
Db	1501	cacaagtgatattgcaggttaa	1522
RESULT 5			
AAA96736			
ID	AAA96736 standard; DNA; 1871 BP.		
AC	AAA96736;		
XX			
DT	19-FEB-2001 (first entry)		
DE	Polynucleotide isolated from lymph node stromal cells of fsn -/- mice.		
XX			
KW	Lymph node stromal cell; fsn -/- mice; inflammatory disorder;		
KW	immune system disorder; cancer; viral infection; HIV infection;		
KW	blood vessel growth; tumour necrosis factor disorder; arthritis;		
KW	inflammatory bowel disease; fibroblast growth factor-mediated disorder;		
KW	cardiac failure; ss.		
XX			
OS	Mus sp.		
XX			
FH	Key Location/Qualifiers		
FT	CDS	64..1575	
FT	/tag= a		
PN	WO200058463-A1.		
XX			
PD	05-OCT-2000.		
XX			
PF	18-FEB-2000; 2000MO-NZ00015.		
XX			

XX Sequence 407 BP; 117 A; 83 C; 95 G; 111 T; 1 other;
SQ

Query Match 12.5%; Score 219.4; DB 20; Length 407;
Best Local Similarity 99.5%; Pred. No. 1.6e-46;
Matches 220; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 901 ggttcagagacatccttggatttgaaacctgacgggcaaatcagagcttccctccatgg 960
|||||
DB 183 ggcttcagacagatcccttggatttgaaacctgacgggcaaatcagagcttccctccatgg 242
QY 961 cagtcgtcaatgagagtgtagaaccaagttcactgtctcctcgccaagcccgacttcag 1020
|||||
DB 243 cagtcgtcaatgagagtgtagaaccaagttcactgtctcctcgccaagcccgacttcag 302
QY 1021 gaccacagcccatcatggttcctcgcgacagtagcaaacaccacacacagagatgg 1080
|||||
DB 303 gaccacagcccatcatggttcctcgcgacagtagcaaacaccacacacagagatgg 362
QY 1081 ctggagatcgatttgggggagaaaagaaataacaggaat 1121
|||||
DB 363 ctggagatcgatttgggggagaaaagaaataacaggaat 403

RESULT 8

AAZ51872
ID AAZ51872 standard; cDNA; 2046 BP.

AC AAZ51872;

DT 04-JUL-2000 (first entry)

XX Human Factor 8 Homologue cDNA.

XX Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;

KW cerebroprotective; therapeutic; coagulation related disorder;

KW haemophilia; stroke; screening; ss.

XX Homo sapiens.

FT key Location/Qualifiers
FT CDS 19..2028
FT /tag= a
FT /product= "Factor 8 Homologue"

PN MO200012532-A1.

XX 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.

XX 31-AUG-1998; 98US-0098521.

XX (ELIL) LILLY & CO ELI.

XX Rostock PRJ, Su W, LI XM;

XX MPI; 2000-256580/22.

XX P-PSDB; AAT70339.

PT Factor 8 homolog polypeptides and nucleic acids encoding them for
treating coagulation related disorders such as hemophilia and stroke -

PS Claim 1; Page 61-64; 68pp; English.

XX The present sequence is a cDNA encoding human Factor 8 homologue (F8H)
CC which is a coagulation cofactor. The protein is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies

CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
of the F8H.

XX Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;
SQ

Query Match 11.7%; Score 205.6; DB 21; Length 2046;
Best Local Similarity 54.2%; Pred. No. 1.2e-42;
Matches 477; Conservative 0; Mismatches 379; Indels 24; Gaps 2;

QY 517 agtgaagaaacagtccttgagagtgatggaatccatcttctggccggggttttctg 576
|||||
DB 187 aatgaataacatctgctgtcatagtggaatccatcttctggacggggttttctg 246
QY 577 acctatgacagagacatccagatttaataacatgtttggaacagctagccatcat 636
|||||
DB 247 tcatctctgttatagataaacaagatcctaattactgtttggacactgcatccatctt 306
QY 637 ttgaagaacagaatatacagaatattcggccagctggttgaagacgtacagagacatt 696
|||||
DB 307 ttgaacctgagttcagtaagtaactgcccagctggttctcctcttcttgagata 366
QY 697 tctgggaatagtagatagatatagatacctcttatttctgcaaaagctgcacatcat 756
|||||
DB 367 tctggaaacatctccatcagatataagatctcctgcacatctgttcagtcgggtgcat 426
QY 757 gcaagaataatctgtcatgaactaagtgtagcagatcagtgcttcagcgcgaaggag 816
|||||
DB 427 gcaagagtagtgtaacaacagcttgtagcgcgaactcagtgcttaagtaagagatc 486
QY 817 agtcgatatgaaggaatctcgcccaatgtgttcttgaaggaatgttccctgtacac 876
|||||
DB 487 cccatattgaagattcttctgtcttaacaacgtcacatctgtgtgttggaactatcata 546
QY 877 aagcgattctgttaccctccatggttgcagcagatccttgagtttgaacctgcagcg 936
|||||
DB 547 agtccttttaactttaagaacagtgatgtagtaaacacgtgggagatgtggtgtgtg 606
QY 937 -----caaatcagagcttctcctcatagtcagtcgtgcatatgagatggagac 984
|||||
DB 607 atcgcgagatccccaataacagcatcatctgtgtgagtgtagtcacacacagggcaa 666
QY 985 caagttcactgtctccctggccaaagcccgacttcagagcacaagcccatctgggtctg 1044
|||||
DB 667 gagaacagtttgaaaacccaataagccagctgaaaacacctggacccgttggtgtgt 726
QY 1045 ggcgacagtagcaacaacacacaacacagagtggtgagatcgatttggggagaaa 1104
|||||
DB 727 ttgccaactgataatc-----agtgttaacaatgatattgaataagaaa 774
QY 1105 aagaataatacagaatattgagccacagatctacacagtcgaacttcaactttatgt 1164
|||||
DB 775 aagaataatacagcattataaacactgataccactgtgtagcacaacttactatgt 834
QY 1165 aagagtttctgtatgaacttcaaaaacataattcttaagtgaagactataaaggaatt 1224
|||||
DB 835 tctgcttaagatctcctgacagtgatgtggcagaatgactgttlaacaggaagcct 894
QY 1225 gtgaataatgaagaagagtggttcaggttaacttaacttcgggacccagtggaaac 1284
|||||
DB 895 ggtgtggagacagaataagatatattcaagaaacagaagtatnaccagagatgtgtgta 954
QY 1285 aattcatcctccatcctggtgcagatatgtgcgggtgttcccccagacatggcaccag 1344
|||||
DB 955 aacttttggccaacatatttgacagtttatttagagtaagtaacttcaacaaatgtagcag 1014
QY 1345 aggatagccttgaaagtgaggtcattgtgtccagata 1384
|||||
DB 1015 aaattgcataagaaatggagctgtcgtgattcagttta 1054


```
Db 6386 aatgccagagcgtgtgaatctgtgcagccagccagcaaacacacataagc-----ag 6439
QY 1078 tggcttgagatcgatttggggggaagaaagaaataacaggaattagaccacagatct 1137
Db 6440 tggcttgagatcgatttggggggaagaaagaaataacaggaattagaccacagatct 6499
QY 1138 acacagtcgaactcaactttaaagagtttgatgaacttgcacaaacataat 1197
Db 6500 aagctctgtcctctgtgaatgtatgaagagccatccatccatcagtcgaagcagga 6559
QY 1198 tctaagtggaaagacctaaaggaattgttaataatgaagaaagggtgttcagggtaac 1257
Db 6560 gtggaatggaaacccatacagcgtgaatccctccatgtgtgacaagatttttgaaggaaat 6619
QY 1258 tctaacttcgggacccagtcgcaaaacaaatctccctccatcgtgacagatattgtg 1317
Db 6620 actaatccaaaagacatgtgaagaaacttttcaaccccccaatcattccaggtttatc 6679
QY 1318 cgggtgtcccccagacatgacacagagataagccttgaaggtgagcctcattgtgtgc 1377
Db 6680 cgtgtcattccctaaacatggaatcaaaagatattgcacttcgcttggaactcttggctgt 6739
QY 1378 cagatt 1383
Db 6740 gatatt 6745
```

RESULT 11

```
AAAT03920
ID AAAT03920 standard; cDNA; 6909 BP.
XX AC AAAT03920;
XX DT 20-DEC-1995 (first entry)
XX DE Human Factor-V wild-type cDNA.
XX KW Factor-V; thrombosis; thrombophilia; diagnosis; anticoagulant;
XX KM activated protein-C; APC; homozygosity; heterozygosity; ss.
XX OS Homo sapiens.
XX PN M09521938-A1.
XX PD 17-AUG-1995.
XX PF 14-FEB-1995; 95MO-EP00553.
XX PR 14-FEB-1994; 94EP-0200377.
XX PA (UYLE-) RIJKSUNIV LEIDEN.
XX PI Bertina RM, Reitsma PH;
XX DR WPI; 1995-293134/38.
XX PT Screening for genetic defect associated with thrombosis and/or poor
XX PT anticoagulant response to activated protein C - useful to determine
XX PT homozygosity or heterozygosity for a mutation in Factor V, Va, VIII
XX PT or VIIIa.
XX PS Claim 16; Page 60-65; 98pp; English.
XX CC The probes/primers given in AAAT03921-31 are used to isolate and/or
XX CC amplify and/or detect mutations in human Factor-V DNA (AAAT03920),
XX CC esp. in DNA encoding the APC binding and/or cleavage site.
XX SQ Sequence 6909 BP; 2096 A; 1700 C; 1430 G; 1683 T; 0 other;
```

Query Match 5.3%; Score 92.8; DB 16; Length 6909;
Best Local Similarity 51.0%; Pred. No. 2.3e-13;

```
Matches 248; Conservative 0; Mismatches 232; Indels 6; Gaps 1;
QY 898 aatggttgacagacatctcttgagtttgaacctgacgggcaaatcagagcttcttccca 957
Db 6280 aatgagatgttccacacocccctgggtatggaataatggaataagaaacaagcaaatcaca 6339
QY 958 tggcagtcggtcatatgaggttgagaccaaagttcattctgtctcctgccaagccgactt 1017
Db 6340 gcttctctttaaagaattcttgggtggagattacttggaaaccttccggtgcccgtctg 6399
QY 1018 caggaccaaagcccatcatatggtctcggcgacagtagcaacaacacacacacagagag 1077
Db 6400 aatgccagggagcgtgtgaatgtccttgcaagccaaagcacaacacataagc-----ag 6453
QY 1078 tggcttgagatcgatttggggggaagaaagaaataacaggaattagaccacagatct 1137
Db 6454 tggctgaatattgtactactcaagatcaagaagataacgcaatttaacacagggctgc 6513
QY 1138 acacagtcgaactcaactttaaagagtttgatgaacttgcacaaacataat 1197
Db 6514 aagctctgtcctctgaatgtatgaagagccatccatccatcagtcgaagcagga 6573
QY 1198 tctaagtggaaagacctaaaggaattgttaataatgaagaaagggtgttcagggtaac 1257
Db 6574 gtggaatggaaacccatacagcgtgaatccctccatgtgtgacaagattttgaaggaaat 6633
QY 1258 tctaacttcgggacccagtcgcaaaacaaatctccctccatcgtgacagatattgtg 1317
Db 6634 actaatccaaaagacatgtgaagaaacttttcaaccccccaatcattccaggtttatc 6693
QY 1318 cgggtgtcccccagacatgacacagagataagccttgaaggtgagcctcattgtgtgc 1377
Db 6694 cgtgtcattccctaaacatggaatcaaaagatattgcacttcgcttggaactcttggctgt 6753
QY 1378 cagatt 1383
Db 6754 gatatt 6759
```

RESULT 12

```
AAA60448
ID AAA60448 standard; cDNA; 6909 BP.
XX AC AAA60448;
XX DT 09-OCT-2000 (first entry)
XX DE Human factor V cDNA sequence SEQ ID NO:26.
XX KW Factor V; FV; activated protein C; APC; anticoagulant;
XX KM activated protein C resistant factor V; thrombosis; screening;
XX KM thrombophilia; ds.
XX OS Homo sapiens.
XX PN US606778-A.
XX PD 23-MAY-2000.
XX PF 06-NOV-1996; 96US-0746111.
XX PR 06-NOV-1996; 96US-0746111.
XX PA (UNMI ) UNIV MICHIGAN.
XX PI Ginsburg D, Cui J;
XX DR WPI; 2000-410682/35.
XX PT New transgenic mice expressing activated protein C resistant factor V
XX PT and factor V null transgenic mice useful for screening anticoagulants,
XX PT as models for human thrombophilia and as models for testing in utero
XX PT gene therapy protocols -
```


Query Match	5.38;	Score 92.8;	DB 17;	Length 6925;
Best Local Similarity	51.08;	Pred. No. 2.3e-13;		
Matches 248; Conservative	0;	Mismatches 232;	Indels 6;	Gaps 1;

QY 988 aatggttcgacagatgacttggatgtttggaactgcacgagcaaatcataggtctctccca 957
DB 6280 aatggaaatgttccaaacccctgggtaacggaaaatggaaatatgaataacaaacaatccca 6339
QY 958 tggagatcgggtcaatggagatggagaccaaagtccatgtgtctctgcgcgaagcccgact 1017
DB 6340 gctcttcgttttaagaatctctggtaggatacttgaatggaaaccttcgcgcgcgctg 6399
QY 1018 caggaaccaaagcccatcatatgagtgcttggggcgagatgagaaataacaaacaaacaaagag 1077
DB 6400 aatggcccaaggagacgtgtgaaatccctggcaagcaagcaacaacaataaagc-----ag 6453
QY 1078 tgggttcgagatcgaatttggggggaagaaagaaataacacgagatlagagaccacagatct 1137
DB 6454 tggctgagaattgtctactacccaagatccaagaagataagcagcatataacacagggctgc 6513
QY 1138 acacagatcgcaattccaacttttatgttaagaggttttgatgaacttcaaaaacatat 1197
DB 6514 aagctctgtctcttgaaatgtagtgaagaaggtataccatccactaaacgtgaagagga 6573
QY 1198 tctaagtcgaagacctataaaggaatttggaaataatgaaagaaagtglttccaggttac 1257
DB 6574 gtggaaatggaaacatcatcagcgctggaatctccatctcatgvtggacaagatgtttgaagaaat 6633
QY 1258 tctaactttcgggagcccaagtcgcaaaaacaaatttatctccctcccatcgttggccaagatatgtg 1317
DB 6634 actaataccaaaaggaacatgtaagaagaactttttaaaccctcccaatcatctccaggttacc 6693
QY 1318 cgggtgtgtcccccgaagatatgcacacgagatatgaccttgaagtggagctcatgttgc 1377
DB 6694 cgtgtcatctcctaaacacatggaatcaaaagtattaaacacttcgcctggaactcttggcgt 6753
QY 1378 cagatt 1383
DB 6754 gatat 6759

Search completed: July 9, 2002, 01:12:07
Job time: 5709 sec

Ultra

691344

=> fil reg
COST IN U.S. DOLLARS

SINCE FILE
ENTRY
0.38

TOTAL
SESSION
2.90

FULL ESTIMATED COST

FILE 'REGISTRY' ENTERED AT 15:46:21 ON 09 JUL 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 American Chemical Society (ACS)

STRUCTURE FILE UPDATES: 8 JUL 2002 HIGHEST RN 437701-77-4
DICTIONARY FILE UPDATES: 8 JUL 2002 HIGHEST RN 437701-77-4

TSCA INFORMATION NOW CURRENT THROUGH January 7, 2002

Please note that search-term pricing does apply when
conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Calculated physical property data is now available. See HELP PROPERTIES
for more information. See STNote 27, Searching Properties in the CAS
Registry File, for complete details:
<http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf>

=> e animal cub domain protein/cn 5

E1 1 ANIMAL BLACK/CN
E2 1 ANIMAL BLOOD, DENATURED/CN
E3 0 --> ANIMAL CUB DOMAIN PROTEIN/CN
E4 1 ANIMAL EXTRACTS, LIVER/CN
E5 1 ANIMAL FAT OLEINS/CN

=> e cub domain/cn 5

E1 1 CUAUII/CN
E2 1 CUAUTHEMONE/CN
E3 0 --> CUB DOMAIN/CN
E4 1 CUB DOMAIN-CONTAINING PROTEIN 1 (HUMAN A549/CALU 6 CELL GENE
CDCP1 PRECURSOR)/CN
E5 1 CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA)/CN

=> e

E6 1 CUB2/CN
E7 1 CUBA BLACK/CN
E8 1 CUBA ORANGE/CN
E9 1 CUBA6/CN
E10 1 CUBANA BROWN/CN
E11 1 CUBANA BROWN R/CN
E12 1 CUBANE/CN
E13 1 CUBANE RADICAL CATION/CN
E14 1 CUBANE-1,2,4,7-TETRAAMMONIUM DINITRAMIDE/CN
E15 1 CUBANE-1,4-DIAMMONIUM DINITRAMIDE/CN
E16 1 CUBANE-1,4-DICARBOXYLIC ACID-1,4-PHENYLENEDIAMINE COPOLYMER/
CN
E17 1 CUBANE-1,4-DICARBOXYLIC ACID-1,4-PHENYLENEDIAMINE COPOLYMER,
SRU/CN

=> s e4-5

1 "CUB DOMAIN-CONTAINING PROTEIN 1 (HUMAN A549/CALU 6 CELL GENE
CDCP1 PRECURSOR)"/CN
1 "CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA)"/CN
L1 2 ("CUB DOMAIN-CONTAINING PROTEIN 1 (HUMAN A549/CALU 6 CELL GENE
CDCP1 PRECURSOR)"/CN OR "CUB-DOMAIN-CONTAINING PROTEIN RP43

Searched by: Mary Hale 308-4258 CM-1 1E01

1. 2. 3.

4.

5.

(RIFTIA PACHYPTILA)"/CN)

=> e coagulation factors v/cn 5

E1	1	COAGULATION FACTOR XIA/CN
E2	1	COAGULATION FACTORS (BLOOD)/CN
E3	0 -->	COAGULATION FACTORS V/CN
E4	1	COAGULEN/CN
E5	3	COAGULIN/CN

=> e coagulation factors xiii/cn 5

E1	1	COAGULATION FACTOR XIA/CN
E2	1	COAGULATION FACTORS (BLOOD)/CN
E3	0 -->	COAGULATION FACTORS XIII/CN
E4	1	COAGULEN/CN
E5	3	COAGULIN/CN

> => e milk fat globule egf factor 8/cn 5

E1	1	MILK BUNDLE-1 PROTEIN (SYNTHETIC)/CN
E2	1	MILK FAT GLOBULE ANTIGEN NP5 (HUMAN FRAGMENT)/CN
E3	0 -->	MILK FAT GLOBULE EGF FACTOR 8/CN
E4	1	MILK GLASS/CN
E5	1	MILK GROWTH FACTOR/CN

=> e transcriptional repressor/cn 5

E1	1	TRANSCRIPTIONAL REGULATORY PROTEIN (YERSINIA PESTIS STRAIN CO 92 GENE YPO1938)/CN
E2	1	TRANSCRIPTIONAL REPRESSION OF HYC AND HYP OPERONS (ESCHERICHIA COLI O157:H7 STRAIN EDL933 GENE HYCA)/CN
E3	0 -->	TRANSCRIPTIONAL REPRESSOR/CN
E4	1	TRANSCRIPTIONAL REPRESSOR (AEDES AEGYPTI STRAIN UGAL REDUCED)/CN
E5	1	TRANSCRIPTIONAL REPRESSOR (AGROBACTERIUM TUMEFACIENS STRAIN C58 GENE ACCR)/CN

=> e

E6	1	TRANSCRIPTIONAL REPRESSOR (AGROBACTERIUM TUMEFACIENS STRAIN C58 GENE TRAM)/CN
E7	1	TRANSCRIPTIONAL REPRESSOR (ARABIDOPSIS THALIANA CLONE G11C3 GENE RPD3A)/CN
E8	1	TRANSCRIPTIONAL REPRESSOR (ASPERGILLUS ORYZAE STRAIN TK3 GENE CREA)/CN
E9	1	TRANSCRIPTIONAL REPRESSOR (BACILLUS HALODURANS STRAIN C-125 GENE BH0826)/CN
E10	1	TRANSCRIPTIONAL REPRESSOR (BACILLUS HALODURANS STRAIN C-125 GENE BH0847)/CN
E11	1	TRANSCRIPTIONAL REPRESSOR (BACTERIOPHAGE .PHI.CH1)/CN
E12	1	TRANSCRIPTIONAL REPRESSOR (BETA-GALACTOSIDASE GENE) (BACILLUS HALODURANS STRAIN C-125 GENE LACR)/CN
E13	1	TRANSCRIPTIONAL REPRESSOR (BRUCELLA MELITENSIS STRAIN 16M GENE BMEI1046)/CN
E14	1	TRANSCRIPTIONAL REPRESSOR (BRUCELLA MELITENSIS STRAIN 2308 GENE ERID)/CN
E15	1	TRANSCRIPTIONAL REPRESSOR (BUTTERFLY GENE EXTRA SEX COMBS)/CN
E16	1	TRANSCRIPTIONAL REPRESSOR (CANDIDA ALBICANS CLONE P371 GENE TUP1 REDUCED)/CN
E17	1	TRANSCRIPTIONAL REPRESSOR (CANDIDA UTILIS STRAIN NRRL Y1084 GENE MIG1)/CN

=> s transcriptional repressor ?/cn

L2 171 TRANSCRIPTIONAL REPRESSOR ?/CN


```

=> e "ae-binding protein-1"/cn 5
E1      1      AE-100/CN
E2      1      AE-3/CN
E3      0 --> AE-BINDING PROTEIN-1/CN
E4      1      AE-CELLULOSE/CN
E5      1      AE-F 117233/CN

=> e neurophilins 1/cn 5
E1      1      NEUROPEPTIDE, SUBOESOPHAGEAL GANGLION (BOMBYX MORI BOMPBAN G
AMMA-)/CN
E2      1      NEUROPEPTIDE-GLUTAMIC ACID-ISOLEUCINE/CN
E3      0 --> NEUROPHILINS 1/CN
E4      1      NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-1 REDUCED)/CN
E5      1      NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-4 REDUCED)/CN

=> e neurophilins 2/cn 5
E1      1      NEUROPEPTIDE, SUBOESOPHAGEAL GANGLION (BOMBYX MORI BOMPBAN G
AMMA-)/CN
E2      1      NEUROPEPTIDE-GLUTAMIC ACID-ISOLEUCINE/CN
E3      0 --> NEUROPHILINS 2/CN
E4      1      NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-1 REDUCED)/CN
E5      1      NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-4 REDUCED)/CN

=> e discoidin domain/cn 5
E1      1      DISCOGUATTINE/CN
E2      1      DISCOIDIC ACID/CN
E3      0 --> DISCOIDIN DOMAIN/CN
E4      1      DISCOIDIN DOMAIN RECEPTOR 1 KINASE/CN
E5      1      DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYM
E DDR1D C-TERMINAL FRAGMENT)/CN

=> e
E6      1      DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYM
E DDR1E FRAGMENT)/CN
E7      1      DISCOIDIN DOMAIN RECEPTOR 2 KINASE/CN
E8      1      DISCOIDIN DOMAIN RECEPTOR KINASE (HUMAN GENE CAK ISOENZYME C
AK II)/CN
E9      1      DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
GENE CAK ISOENZYME CAK I)/CN
E10     1      DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
GENE CAK ISOENZYME CAK II)/CN
E11     1      DISCOIDIN DOMAIN RECEPTOR TYROSINE KINASE/CN
E12     1      DISCOIDIN DP1 (DICTYOSTELIUM DISCOIDEUM STRAIN AX2)/CN
E13     1      DISCOIDIN I-A (DICTYOSTELIUM DISCOIDEUM REDUCED)/CN
E14     1      DISCOIDIN I-C (DICTYOSTELIUM DISCOIDEUM REDUCED)/CN
E15     1      DISCOIDIN II (DICTYOSTELIUM DISCOIDEUM CLONE D2)/CN
E16     1      DISCOKIOLIDE A/CN
E17     1      DISCOKIOLIDE B/CN

=> s e4-11
1 "DISCOIDIN DOMAIN RECEPTOR 1 KINASE"/CN
1 "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME
DDR1D C-TERMINAL FRAGMENT)"/CN
1 "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME
DDR1E FRAGMENT)"/CN
1 "DISCOIDIN DOMAIN RECEPTOR 2 KINASE"/CN
1 "DISCOIDIN DOMAIN RECEPTOR KINASE (HUMAN GENE CAK ISOENZYME CAK
II)"/CN
1 "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
GENE CAK ISOENZYME CAK I)"/CN
1 "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
GENE CAK ISOENZYME CAK II)"/CN

```


L3 1 "DISCOIDIN DOMAIN RECEPTOR TYROSINE KINASE"/CN
 8 ("DISCOIDIN DOMAIN RECEPTOR 1 KINASE"/CN OR "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME DDR1D C-TERMINAL FRAGMENT)"/CN OR "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME DDR1E FRAGMENT)"/CN OR "DISCOIDIN DOMAIN RECEPTOR 2 KINASE"/CN OR "DISCOIDIN DOMAIN RECEPTOR KINASE (HUMAN GENE CAK ISOENZYME CAK II)"/CN OR "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL GENE CAK ISOENZYME CAK I)"/CN OR "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL GENE CAK ISOENZYME CAK II)"/CN OR "DISCOIDIN DOMAIN RECEPTOR TYROSINE KINASE"/CN)

=> e nhp/cn 5

E1 1 NHJ 44/CN
 E2 1 NHMO/CN
 E3 0 --> NHP/CN
 E4 1 NHP2/RS6-LIKE PROTEIN (TRYPANOSOMA BRUCEI STRAIN 427 GENE NHP2)/CN
 E5 1 NHPTT/CN

=> fil medl,hcap,biosis,embase,jicst,wpids
 COST IN U.S. DOLLARS

SINCE FILE	TOTAL
ENTRY	SESSION
46.28	49.18

FULL ESTIMATED COST

FILE 'MEDLINE' ENTERED AT 15:49:39 ON 09 JUL 2002

FILE 'HCAPLUS' ENTERED AT 15:49:39 ON 09 JUL 2002
 USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
 PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
 COPYRIGHT (C) 2002 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'BIOSIS' ENTERED AT 15:49:39 ON 09 JUL 2002
 COPYRIGHT (C) 2002 BIOLOGICAL ABSTRACTS INC. (R)

FILE 'EMBASE' ENTERED AT 15:49:39 ON 09 JUL 2002
 COPYRIGHT (C) 2002 Elsevier Science B.V. All rights reserved.

FILE 'JICST-EPLUS' ENTERED AT 15:49:39 ON 09 JUL 2002
 COPYRIGHT (C) 2002 Japan Science and Technology Corporation (JST)

FILE 'WPIDS' ENTERED AT 15:49:39 ON 09 JUL 2002
 COPYRIGHT (C) 2002 THOMSON DERWENT

=> s (l1 or cub domain protein) and (coagulat? factor? or blood coagulation or milk fat globule egf factor 8 or mfg human milk or transcriptional repressor or crtr-1 protein or ae-binding protein 1 or neurophilin?)

L4 0 FILE MEDLINE
 L5 1 FILE HCAPLUS
 L6 0 FILE BIOSIS
 L7 0 FILE EMBASE
 L8 0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA/CN'

L9 1 FILE WPIDS

TOTAL FOR ALL FILES

L10 2 (L1 OR CUB DOMAIN PROTEIN) AND (COAGULAT? FACTOR? OR BLOOD COAGU

Searched by: Mary Hale 308-4258 CM-1 1E01

LATION OR MILK FAT GLOBULE EGF FACTOR 8 OR MFG HUMAN MILK OR
TRANSCRIPTIONAL REPRESSOR OR CRTR-1 PROTEIN OR AE-BINDING PROTEI
N 1 OR NEUROPHILIN?)

=> dup rem l10
PROCESSING COMPLETED FOR L10
L11 1 DUP REM L10 (1 DUPLICATE REMOVED)

=> d cbib abs

L11 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 1
2001:300872 Document No. 134:321614 Protein and cDNA sequences of a novel
human CUB-domain containing protein and uses thereof in diagnosis, therapy
and drug screening. Donoho, Gregory; Turner, C. Alexander, Jr.; Nehls,
Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon
Genetics Incorporated, USA). PCT Int. Appl. WO 2001029219 A1 20010426, 33
pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR,
BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD,
SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM,
CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT,
SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US28798
20001008. PRIORITY: US 1999-PV160285 19991019; US 2000-PV183583 20000218.

AB This invention provides protein and cDNA sequences for newly identified
human proteins, designated NHPs, which shares structural similarity with
animal **CUB domain proteins**,
coagulation factors V and XIII, **milk**
fat globule-EGF factor 8,
5 transcriptional repressor AE-
binding protein-1, and **neuropilins 1 and 2**
(which, like the presently described protein, contain both CUB and
discoidin domains). While NHP shares sequence homol. with other
CUB-domain contg. proteins, its primary sequence is unique. Its
expression is detected in human cell lines, and human prostate, pituitary,
fetal brain, brain, thymus, spleen, lymph node, trachea, kidney, fetal
liver, thyroid, adrenal gland, salivary gland, stomach, small intestine,
colon, muscle, heart, mammary gland, adipose, skin, esophagus, bladder,
cervix, rectum, and testis cells. In one embodiment, the invention
relates to diagnostic assays for detecting diseases assocd. with
inappropriate NHP activity or levels. Also disclosed are methods for
utilizing NHP in drug screening assays and in therapy directed against
diseases assocd. with inappropriate NHP activity or levels.

=> s (l1 or cub domain protein) and (discoidin domain or l3 or discoidin receptor
or ddr2 gene product?)
L12 0 FILE MEDLINE
L13 1 FILE HCAPLUS
L14 0 FILE BIOSIS
L15 0 FILE EMBASE
L16 0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

Searched by: Mary Hale 308-4258 CM-1 1E01

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

L17 0 FILE WPIDS

TOTAL FOR ALL FILES

L18 1 (L1 OR CUB DOMAIN PROTEIN) AND (DISCOIDIN DOMAIN OR L3 OR DISCOI
DIN RECEPTOR OR DDR2 GENE PRODUCT?)

=> s l18 not l10

L19 0 FILE MEDLINE
L20 0 FILE HCAPLUS
L21 0 FILE BIOSIS
L22 0 FILE EMBASE
L23 0 FILE JICST-EPLUS
L24 0 FILE WPIDS

TOTAL FOR ALL FILES

L25 0 L18 NOT L10

=> (l1 or cub domain or extracellular domain or ecd) and (novel human protein? or
nhps)

(L1 IS NOT A RECOGNIZED COMMAND

The previous command name entered was not recognized by the system.

For a list of commands available to you in the current file, enter

"HELP COMMANDS" at an arrow prompt (=>).

=> s (l1 or cub domain or extracellular domain or ecd) and (novel human protein? or
nhps)

L26 1 FILE MEDLINE
L27 7 FILE HCAPLUS
L28 1 FILE BIOSIS
L29 1 FILE EMBASE
L30 0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'

L31 1 FILE WPIDS

TOTAL FOR ALL FILES

L32 11 (L1 OR CUB DOMAIN OR EXTRACELLULAR DOMAIN OR ECD) AND (NOVEL
HUMAN PROTEIN? OR NHPS)

=> s l32 not (l10 or l18)

L33 1 FILE MEDLINE
L34 6 FILE HCAPLUS
L35 1 FILE BIOSIS
L36 1 FILE EMBASE
L37 0 FILE JICST-EPLUS
L38 0 FILE WPIDS

Searched by: Mary Hale 308-4258 CM-1 1E01

TOTAL FOR ALL FILES

L39 9 L32 NOT (L10 OR L18)

=> dup rem l39

PROCESSING COMPLETED FOR L39

L40 6 DUP REM L39 (3 DUPLICATES REMOVED)

=> d cbib abs 1-6

L40 ANSWER 1 OF 6 HCAPLUS COPYRIGHT 2002 ACS

2002:107557 Document No. 136:162371 Cloning and characterization of
novel human protein kinase family members

32374 and 18431 and their therapeutic uses. Meyers, Rachel;
Kapeller-Libermann, Rosana; Silos-Santiago, Immaculada (Millennium
Pharmaceuticals, Inc., USA). PCT Int. Appl. WO 2002010401 A2 20020207,
141 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG,
BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT,
RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN,
YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF,
CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML,
MR, NE, NL, PT, SE, SN, TD, TG, TR. (English). CODEN: PIXXD2.
APPLICATION: WO 2001-US23653 20010727. PRIORITY: US 2000-PV221543
20000728.

AB The invention provides isolated nucleic acids mols., designated 32374 or
18431 nucleic acid mols., which encode novel protein kinase family
members. The invention also provides antisense nucleic acid mols.,
recombinant expression vectors contg. 32374 or 18431 nucleic acid mols.,
host cells into which the expression vectors have been introduced, and
nonhuman transgenic animals in which a 32374 or 18431 gene has been
introduced or disrupted. Their putative function domains are analyzed and
their gene expression profiles are provided. The invention still further
provides isolated 32374 or 18431 proteins, fusion proteins, antigenic
peptides and anti-32374 or -18431 antibodies. Diagnostic methods
utilizing compns. of the invention are also provided.

L40 ANSWER 2 OF 6 HCAPLUS COPYRIGHT 2002 ACS

2002:10647 Document No. 136:80918 Protein and cDNA sequences of a novel
human **CUB domain**-containing protein sequence homolog
and uses thereof. Kapeller-Libermann, Rosana (Millennium Pharmaceuticals,
Inc., USA). PCT Int. Appl. WO 2002000843 A2 20020103, 107 pp. DESIGNATED
STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH,
CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,
LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY,
DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE,
SN, TD, TG, TR. (English). CODEN: PIXXD2. APPLICATION: WO 2001-US20055
20010621. PRIORITY: US 2000-PV213963 20000623.

AB The invention provides protein and cDNA sequences of a **novel**
human protein, designated 56739, which has sequence
homol. with CUB family members. The invention also provides antisense
nucleic acid mols., recombinant expression vectors contg. 56739 nucleic
acid mols., host cells into which the expression vectors have been
introduced, and nonhuman transgenic animals in which a 56739 gene has been
introduced or disrupted. The invention still further provides isolated
56739 proteins, fusion proteins, antigenic peptides and anti-56739
antibodies. Diagnostic methods utilizing compns. of the invention are
also provided.

L40 ANSWER 3 OF 6 HCAPLUS COPYRIGHT 2002 ACS

2001:526099 Document No. 135:117953 cDNA and protein sequence of

novel human protein TANGO416 and TANGO457 and

their uses in drug screening, therapeutics and diagnosis. Fraser, Christopher C.; Kirst, Susan J. (Millennium Pharmaceuticals, Inc., USA).

PCT Int. Appl. WO 2001051514 A1 20010719, 217 pp. DESIGNATED STATES: W:

AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG, TR.

(English). CODEN: PIXXD2. APPLICATION: WO 2001-US485 20010105.

PRIORITY: US 2000-479249 20000107; US 2000-559497 20000427.

AB The invention provides cDNA and protein sequence of two **novel human protein** TANGO416 and TANGO457. The protein

TANGO416 was cloned from human fetal spleen and has high homol. with murine protocadherin. Both proteins are transmembrane proteins and contg. extracellular, transmembrane and intracellular domain. These nucleic acid and proteins are useful for diagnosis, prevention, and therapy of a no. of human and other animal disorders. The invention also provides antisense nucleic acid mols., expression vectors contg. the nucleic acid mols. of the invention, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a nucleic acid mol. of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening, and therapeutic methods using comps. of the invention are also provided. The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes.

L40 ANSWER 4 OF 6 HCAPLUS COPYRIGHT 2002 ACS

2001:300759 Document No. 134:321609 Protein and cDNA sequences of human

secretory proteins and uses thereof in diagnosis, therapy and drug

screening. Turner, C. Alexander, Jr.; Donoho, Gregory; Nehls, Michael;

Hilbun, Erin; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics

Incorporated, USA). PCT Int. Appl. WO 2001029084 A2 20010426, 63 pp.

DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ,

CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,

HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,

LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG,

SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY,

KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE,

DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN,

TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US28803

20001018. PRIORITY: US 1999-PV160106 19991018; US 1999-PV162547 19991029.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated **NHPs**, which shares structure motifs typical of human secreted proteins that share structural similarity with animal chordins, NEL protein, and thrombospondin. Also disclosed is another **novel human protein** that shares structural similarity with animal protein that contains **CUB domain**. In one embodiment, the invention relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels. Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels.

L40 ANSWER 5 OF 6 HCAPLUS COPYRIGHT 2002 ACS

1998:608707 Document No. 129:215729 Cloning of cDNA for **novel**

Searched by: Mary Hale 308-4258 CM-1 1E01

human proteins belong to tumor necrosis factor receptor superfamily. Tada, Hideaki; Konishi, Mikio; Fukushima, Daikichi (Ono Pharmaceutical Co., Ltd., Japan). PCT Int. Appl. WO 9838304 A1 19980903, 54 pp. DESIGNATED STATES: W: JP, KR, US; RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE. (Japanese). CODEN: PIXXD2. APPLICATION: WO 1998-JP799 19980226. PRIORITY: JP 1997-43143 19970227.

AB The cDNA for proteins OAF065.alpha. and OAF065.beta. are isolated from human stroma cell line HAS303. OAF065.alpha. consists of 416 and OAF065.beta. of 423 amino acids, with identical N-terminal 1-415 residues. The C-terminal Glu-Ala of OAF065.alpha. is substituted with 8 amino acid residues in OAF065.beta.. Both protein contains a Cys-rich domain, commonly seen in the **extracellular domain** of the TNF receptor superfamily. Claimed are methods of recombinant prepn. of the proteins, (monoclonal) antibody to the proteins, oligonucleotide hybridizing to the cDNA, and therapeutics contg. the protein and/or the antibodies.

L40 ANSWER 6 OF 6 MEDLINE DUPLICATE 1
96394649 Document Number: 96394649. PubMed ID: 8798755. Cloning and characterization of islet cell antigen-related protein-tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen in insulin-dependent diabetes. Cui L; Yu W P; DeAizpurua H J; Schmidli R S; Pallen C J. (Cell Regulation Laboratory, Institute of Molecular and Cell Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511, Republic of Singapore.) JOURNAL OF BIOLOGICAL CHEMISTRY, (1996 Oct 4) 271 (40) 24817-23. Journal code: 2985121R. ISSN: 0021-9258. Pub. country: United States. Language: English.

AB Cloning of the cDNA encoding a **novel human protein-** tyrosine phosphatase (PTP) called islet cell antigen-related PTP (IAR) predicts a receptor-like molecule with an **extracellular domain** of 614 amino acids containing a hydrophobic signal peptide, one potential N-glycosylation site, and an RGDS peptide which is a possible adhesive recognition sequence. The 376-amino acid intracellular region contains a single catalytic domain. Recombinant IAR polypeptide has phosphatase activity. Northern blot analysis shows tissue-specific expression of two IAR transcripts of 5.5 and 3.7 kilobases, which are most abundant in brain and pancreas. The IAR PTP is homologous in its intracellular region to IA-2, a putative PTP that is an insulin-dependent diabetes mellitus (IDDM) autoantigen. IAR is also reactive with IDDM patient sera. IAR and IA-2 may distinguish different populations of IDDM autoantibodies since they identify overlapping but nonidentical sets of IDDM patients. Thus IAR is likely to be an islet cell antigen useful in the preclinical screening of individuals for risk of IDDM.

=> s (l1 or cub domain or extracellular domain or ecd) and (agonist or antagonist)

L41 611 FILE MEDLINE
L42 485 FILE HCAPLUS
L43 408 FILE BIOSIS
L44 359 FILE EMBASE
L45 36 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA/CN'

L46 203 FILE WPIDS

TOTAL FOR ALL FILES

L47 2102 (L1 OR CUB DOMAIN OR EXTRACELLULAR DOMAIN OR ECD) AND (AGONIST

Searched by: Mary Hale 308-4258 CM-1 1E01

OR ANTAGONIST)

=> s 147 and (l3 or discoidin(w) (receptor! or domain))

L48 0 FILE MEDLINE
L49 1 FILE HCAPLUS
L50 0 FILE BIOSIS
L51 0 FILE EMBASE
L52 0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

L53 0 FILE WPIDS

TOTAL FOR ALL FILES

L54 1 L47 AND (L3 OR DISCOIDIN(W) (RECEPTOR! OR DOMAIN))

=> d cbib abs

L54 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2002 ACS

1998:71221 Document No. 128:136859 Soluble ErbB receptor

extracellular domain fusion proteins their uses in
antagonization of growth factors. Fitzpatrick, Vincent Danial; Sliwowski,
Mark; Vandlen, Richard L. (Genentech, Inc., USA). PCT Int. Appl. WO
9802540 A1 19980122, 58 pp. DESIGNATED STATES: W: AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH,
HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG,
MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW:
AT, BE, BF, BJ, CF, CG, CH, CI, CM, DE, DK, ES, FI, FR, GA, GB, GR, IE,
IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN:
PIXXD2. APPLICATION: WO 1997-US11825 19970708. PRIORITY: US 1996-21640
19960712; US 1997-798326 19970210.

AB Sol. heteromultimeric immunoadhesin fusion proteins that contain the
extracellular ligand-binding domains of ErbB receptors oligomerized via
the multimerization domains of Igs are described for use as
antagonists of neuregulins in the control of cell proliferation,
e.g. in cancers. Specifically, the **extracellular**
domains of ErbB2 and ErbB3 or ErbB2 and ErbB4 are used in
combination. The prepn. ErbB/IgG fusion proteins and the characterization
of their heregulin binding properties is reported.

=> s donoho, g?/au,in;s turner, c?/au;s nehls, m?/au,in;s friedrich, g?/au,in
'IN' IS NOT A VALID FIELD CODE

L55 6 FILE MEDLINE
L56 44 FILE HCAPLUS
L57 9 FILE BIOSIS
'IN' IS NOT A VALID FIELD CODE
L58 5 FILE EMBASE
L59 0 FILE JICST-EPLUS

Searched by: Mary Hale 308-4258 CM-1 1E01

L60 36 FILE WPIDS

TOTAL FOR ALL FILES

L61 100 DONOHO, G?/AU, IN

L62 1091 FILE MEDLINE

L63 1428 FILE HCAPLUS

L64 1198 FILE BIOSIS

L65 772 FILE EMBASE

L66 7 FILE JICST-EPLUS

L67 212 FILE WPIDS

TOTAL FOR ALL FILES

L68 4708 TURNER, C?/AU

'IN' IS NOT A VALID FIELD CODE

L69 38 FILE MEDLINE

L70 77 FILE HCAPLUS

L71 42 FILE BIOSIS

'IN' IS NOT A VALID FIELD CODE

L72 30 FILE EMBASE

L73 0 FILE JICST-EPLUS

L74 39 FILE WPIDS

TOTAL FOR ALL FILES

L75 226 NEHLS, M?/AU, IN

'IN' IS NOT A VALID FIELD CODE

L76 158 FILE MEDLINE

L77 337 FILE HCAPLUS

L78 163 FILE BIOSIS

'IN' IS NOT A VALID FIELD CODE

L79 167 FILE EMBASE

L80 1 FILE JICST-EPLUS

L81 105 FILE WPIDS

TOTAL FOR ALL FILES

L82 931 FRIEDRICH, G?/AU, IN

=> s l61 and l68 and l75 and l82

L83 0 FILE MEDLINE

L84 10 FILE HCAPLUS

L85 0 FILE BIOSIS

L86 0 FILE EMBASE

L87 0 FILE JICST-EPLUS

L88 10 FILE WPIDS

TOTAL FOR ALL FILES

L89 20 L61 AND L68 AND L75 AND L82

=> dup rem l89

PROCESSING COMPLETED FOR L89

L90 10 DUP REM L89 (10 DUPLICATES REMOVED)

=> d 1-10 cbib abs

L90 ANSWER 1 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 1
2001:360153 Document No. 134:363350 Characterization and applications of
novel human proteases. Donoho, Gregory; Hilbun, Erin; Turner,

Searched by: Mary Hale 308-4258 CM-1 1E01

C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001034779 A2 20010517, 34 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG, TR. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US31179 20001109. PRIORITY: US 1999-PV165260 19991112.

AB The cDNA sequences and corresponding deduced amino acid sequences of novel human proteases (NHPs) are disclosed. The NHP genes were obtained from a human testis cDNA library using probes and/or primers generated from human gene trapped sequence tags. The described NHPs share significant similarity to a variety of mammalian proteases, and esp. carboxypeptidase A. The NHPs can be used in therapeutic, diagnostic, and pharmacogenomic applications.

L90 ANSWER 2 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 2
2001:360152 Document No. 134:363349 Characterization of novel human ATPase homologs. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Hilbun, Erin; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001034778 A2 20010517, 45 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG, TR. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US31110 20001109. PRIORITY: US 1999-PV164624 19991110.

AB The present invention relates to the discovery, identification, and characterization of novel human proteins which share structural similarity with calcium transporting ATPases and aminophospholipid transporters. The cDNA sequences and the corresponding deduced amino acid sequences of the human ATPase homologs are disclosed. The human ATPase homologs can be used in therapeutic, diagnostic, and pharmacogenomic applications.

L90 ANSWER 3 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 3
2001:338573 Document No. 134:348997 Protein and cDNA sequences of human transport proteins and uses thereof in diagnosis, therapy and drug screening. Harras, Marie; **Donoho, Gregory**; **Turner, C. Alexander, Jr.**; **Nehls, Michael**; **Friedrich, Glenn**; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001032706 A2 20010510, 59 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US29852 20001031. PRIORITY: US 1999-PV163018 19991102.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which share structural similarity with mammalian multi-drug resistance (MDR) proteins and cellular transporters. In one embodiment, the invention relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels.

Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels. Seq ID No. 47 and 48 were claimed, but information was not provided.

L90 ANSWER 4 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 4
2001:300872 Document No. 134:321614 Protein and cDNA sequences of a novel human CUB-domain containing protein and uses thereof in diagnosis, therapy and drug screening. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001029219 A1 20010426, 33 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US28798 20001008. PRIORITY: US 1999-PV160285 19991019; US 2000-PV183583 20000218.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which shares structural similarity with animal CUB domain proteins, coagulation factors V and XIII, milk fat globule-EGF factor 8, 5 transcriptional repressor AE-binding protein-1, and neuropilins 1 and 2 (which, like the presently described protein, contain both CUB and discoidin domains). While NHP shares sequence homol. with other CUB-domain contg. proteins, its primary sequence is unique. Its expression is detected in human cell lines, and human prostate, pituitary, fetal brain, brain, thymus, spleen, lymph node, trachea, kidney, fetal liver, thyroid, adrenal gland, salivary gland, stomach, small intestine, colon, muscle, heart, mammary gland, adipose, skin, esophagus, bladder, cervix, rectum, and testis cells. In one embodiment, the invention relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels. Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels.

L90 ANSWER 5 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 5
2001:247518 Document No. 134:276537 Protein and cDNA sequences of human carboxypeptidases and uses thereof in diagnosis, therapy and drug screening. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001023588 A1 20010405, 36 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26876 20000929. PRIORITY: US 1999-PV156685 19990929.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which shares substantial sequence homol. with animal proteases, esp. carboxypeptidases. While NHP shares sequence homol. with other carboxypeptidases, its primary sequence is unique. Its expression is detected in various human tissues including brain, pituitary, spinal cord, thymus, spleen, lymph node, bone marrow, trachea, lung, kidney, prostate, testis, thyroid, adrenal gland, stomach, small intestine colon, skeletal muscle, uterus, mammary gland, bladder, cervix cells, and gene trapped human cells. In one embodiment, the invention

relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels. Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels.

L90 ANSWER 6 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 6
2001:247510 Document No. 134:261891 Protein and cDNA sequences of human serine/threonine protein kinase and uses thereof in diagnosis, therapy and drug screening. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001023579 A1 20010405, 38 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26621 20000927. PRIORITY: US 1999-PV156511 19990928.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which shares substantial sequence homol. with animal kinases, and more particular serine/threonine protein kinases. While NHP shares sequence homol. with other serine/threonine protein kinases, its primary sequence is unique. Its expression is detected in various human tissues including brain, pituitary, spinal cord, spleen, trachea, kidney, prostate, testis, adrenal gland cells, and gene trapped human cells. In one embodiment, the invention relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels. Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels.

L90 ANSWER 7 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 7
2001:247499 Document No. 134:261870 Protein and cDNA sequences of a novel human organic anion transport protein. **Turner, C. Alexander, Jr.**; **Donoho, Gregory**; Wattler, Frank; **Nehls, Michael**; **Friedrich, Glenn**; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001023566 A1 20010405, 31 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26630 20000927. PRIORITY: US 1999-PV156161 19990927.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which shares substantial sequence homol. with animal org. anion, and more particularly prostaglandin, transporter proteins. While NHP shares sequence homol. with other org. anion transport proteins, its primary sequence is unique. Its expression is detected in various human tissues including brain, pituitary, spinal cord, lymph node, trachea, heart, adipose, skin, pericardium, hypothalamus cells, and gene trapped human cells. In one embodiment, the invention relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels. Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels.

L90 ANSWER 8 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 8

2001:229020 Document No. 134:262848 Cloning and characterization of human endothelin converting enzyme-like proteins. **Donoho, Gregory;** Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001021773 A2 20010329, 31 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26047 20000922. PRIORITY: US 1999-PV156102 19990924; US 2000-PV176689 20000118.

AB Novel human endothelin converting enzyme-like proteins of 255 and 883 amino acids in length and cDNAs encoding these proteins are disclosed. The endothelin converting enzyme-like proteins and encoding nucleic acids can be used in therapeutic, diagnostic, and pharmacogenomic applications.

L90 ANSWER 9 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 9

2001:228919 Document No. 134:247996 Protein and cDNA sequences for a novel human protease inhibitor-like protein NHP and use thereof. **Donoho, Gregory;** Turner, C. Alexander, Jr.; Wattler, Frank; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001021651 A2 20010329, 29 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26048 20000922. PRIORITY: US 1999-PV156101 19990924.

AB The invention provides protein and cDNA sequences for a novel human protease inhibitor-like protein NHP that can be used in therapeutic, diagnostic, and pharmacogenomic applications.

L90 ANSWER 10 OF 10 HCAPLUS COPYRIGHT 2002 ACS DUPLICATE 10

2001:168158 Document No. 134:189017 Protein and cDNA sequences of human calcium dependent protease homologs, and uses thereof in therapy, diagnosis and drug screening. **Donoho, Gregory;** Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001016336 A1 20010308, 39 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US24062 20000901. PRIORITY: US 1999-PV152057 19990902.

AB The invention provides protein and cDNA sequences of a novel class of human calcium dependent protease homologs which share structural similarity with animal calcium dependent proteases, or calpains. The novel calcium dependent protease homolog encoding cDNAs were obtained from human gene trap clones and human cDNA libraries. The invention further relates to the uses of calcium dependent protease homolog in therapy, diagnosis and drug screening.

=> log y

COST IN U.S. DOLLARS	SINCE FILE	TOTAL
	ENTRY	SESSION
FULL ESTIMATED COST	71.52	120.70
DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS)	SINCE FILE	TOTAL
	ENTRY	SESSION
CA SUBSCRIBER PRICE	-10.53	-10.53

STN INTERNATIONAL LOGOFF AT 16:04:39 ON 09 JUL 2002

